GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 13:03:17 ; Search time 3909 Seconds
(without alignments)
10381.773 Million cell updates/sec

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggcacgagccaggaactagg.....aaaaaaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		992	100.0	992	6	AR252529	AR252529 Sequence
2		992	100.0	992	6	AX358896	AX358896 Sequence
3		992	100.0	992	6	AX362389	AX362389 Sequence
4		992	100.0	992	6	AX403335	AX403335 Sequence
5		504.6	50.9	505	6	BD107970	BD107970 EST and e
6		501.6	50.6	502	6	AX588820	AX588820 Sequence
7	c	409.8	41.3	414	11	G60502	G60502 SHGC-33533
8	c	300.4	30.3	154840	2	AC032035	AC032035 Homo sapi
9	c	300.4	30.3	177738	9	AC021593	AC021593 Homo sapi
10	c	300.4	30.3	209751	9	AC087645	AC087645 Homo sapi
11		292.8	29.5	229426	2	AC010532	AC010532 Homo sapi
12		291.2	29.4	154840	2	AC032035	AC032035 Homo sapi
13	c	179.4	18.1	229426	2	AC010532	AC010532 Homo sapi
14		163.4	16.5	116300	10	AL645856	AL645856 Mouse DNA
15		163.4	16.5	197194	2	AL954690	AL954690 Mus muscu
16		160.2	16.1	49532	2	AC100565	AC100565 Mus muscu
17	c	149	15.0	150	11	G43545	G43545 WIAF-2405-S
18	c	149	15.0	150	11	G43546	G43546 WIAF-2407-S
19		127.8	12.9	161043	2	AC143361	AC143361 Macaca mu
20		69.2	7.0	151750	9	AC003666	AC003666 Homo sapi
21		69.2	7.0	208670	9	AC103588	AC103588 Homo sapi
22		62.8	6.3	177921	9	AL353752	AL353752 Human DNA
23	c	61.8	6.2	146883	9	AC018788	AC018788 Homo sapi
24		61	6.1	125020	9	AF429315	AF429315 Homo sapi
25		60.8	6.1	164762	2	AC016177	AC016177 Homo sapi
26	c	60.8	6.1	170898	9	AL353764	AL353764 Human DNA
27		60.4	6.1	129770	9	AC104684	AC104684 Homo sapi
28	c	60.4	6.1	171655	2	AC067919	AC067919 Homo sapi
29		58.8	5.9	124347	9	AC010072	AC010072 Homo sapi
30	c	58.6	5.9	151993	9	AC093574	AC093574 Homo sapi
31		58.6	5.9	194635	9	AC012087	AC012087 Homo sapi
32	c	58.4	5.9	4950	9	HSM130AC2	Z22970 H.sapiens m
33		58.4	5.9	177260	9	AC131206	AC131206 Homo sapi
34		58	5.8	140055	2	AC023003	AC023003 Homo sapi
35		58	5.8	151553	9	AL139342	AL139342 Human DNA
36	c	58	5.8	163085	9	AC019159	AC019159 Homo sapi
37		57.8	5.8	99268	9	AC091178	AC091178 Homo sapi
38		57.8	5.8	119853	9	AL136374	AL136374 Human DNA
39		57.8	5.8	138370	9	AL354713	AL354713 Human DNA
40		57.8	5.8	156857	2	AC025375	AC025375 Homo sapi
41	c	57.8	5.8	161940	2	AC079289	AC079289 Homo sapi
42	c	57.4	5.8	50630	9	AC107054	AC107054 Homo sapi
43		57.4	5.8	145380	9	AL357129	AL357129 Human DNA
44		57.4	5.8	169202	9	AC021021	AC021021 Homo sapi
45	c	57.4	5.8	175947	2	AC145171	AC145171 Homo sapi
46		56.6	5.7	151841	9	AC013408	AC013408 Homo sapi
47		56.6	5.7	173270	9	AC025522	AC025522 Homo sapi
48		56.4	5.7	164405	9	AC016256	AC016256 Homo sapi
49	c	56.4	5.7	166668	2	AC025258	AC025258 Homo sapi
50	c	56	5.6	183778	2	AC011004	AC011004 Homo sapi
51	c	56	5.6	199437	2	AC091575	AC091575 Homo sapi
52		56	5.6	207069	9	AC090922	AC090922 Homo sapi
53	c	56	5.6	254455	2	AC099902	AC099902 Mus muscu
54		55.8	5.6	112389	2	AC073598	AC073598 Homo sapi
55	c	55.8	5.6	189271	9	AL138752	AL138752 Human DNA
56	c	55.4	5.6	214866	9	AC069544	AC069544 Homo sapi
57	c	54.6	5.5	169986	2	AL353589	AL353589 Homo sapi
58	c	54.6	5.5	181356	2	AC069441	AC069441 Homo sapi
59	c	54.6	5.5	183104	9	AL355601	AL355601 Human DNA
60		54.6	5.5	193793	9	AC099780	AC099780 Homo sapi
61	c	54.6	5.5	208026	9	AP006292	AP006292 Homo sapi
62		54.4	5.5	207190	9	AF295017	AF295017 Homo sapi
63	c	54.4	5.5	217898	9	AC008498	AC008498 Homo sapi
64		54.2	5.5	106340	9	AC097104	AC097104 Homo sapi
65	c	54.2	5.5	146017	2	AC027473	AC027473 Homo sapi

c	66	54.2	5.5	151465	2	AC117371
	67	54.2	5.5	154608	2	AC073620
c	68	54	5.4	231746	9	AC011389
	69	53.8	5.4	91076	9	AC073892
	70	53.8	5.4	95615	9	AL138835
c	71	53.8	5.4	159123	9	AC106871
c	72	53.8	5.4	176735	2	AC021271
	73	53.8	5.4	180275	2	AC024482
	74	53.8	5.4	181864	2	AC027460
	75	53.6	5.4	166239	2	AC078965
	76	53.6	5.4	167071	9	AC092380
c	77	53.6	5.4	168822	9	AC007511
	78	53.4	5.4	171100	9	AC092371
	79	53.4	5.4	173548	9	AC092136
	80	53.4	5.4	173905	9	AC092125
c	81	53.4	5.4	182528	9	AC020743
c	82	53.2	5.4	85149	9	AL137879
c	83	53.2	5.4	119814	9	AL606535
	84	53.2	5.4	170124	9	AC067733
	85	53.2	5.4	184121	9	AC016706
	86	53.2	5.4	278772	2	AC067734
c	87	52.8	5.3	81514	9	HSAJ9616
c	88	52.8	5.3	192307	9	AC005971
	89	52.6	5.3	145309	2	AC073414
	90	52.6	5.3	165113	9	AP001032
	91	52.6	5.3	175132	9	AL359641
	92	52.6	5.3	180120	9	AC090857
	93	52.6	5.3	180434	2	AC019137
c	94	52.6	5.3	211325	2	AP002757
c	95	52	5.2	84613	9	AL136140
c	96	52	5.2	136713	9	AC112255
	97	51.6	5.2	159158	9	AL359915
	98	51.6	5.2	185652	9	AC007312
c	99	51.4	5.2	100000	9	AB020873
	100	51.4	5.2	100791	9	CNS01RI8
c	101	51.4	5.2	125196	9	AL731897
c	102	51.4	5.2	165755	9	AP006210
c	103	51.4	5.2	175726	9	AL691420
	104	51.2	5.2	141447	9	AC016204
	105	51.2	5.2	180233	9	AC011767
c	106	51	5.1	64791	9	AL353649
c	107	51	5.1	179207	2	AC009900
c	108	50.4	5.1	116236	9	AL354877
	109	50.4	5.1	146463	9	AC114971
	110	50.4	5.1	150013	2	AC010780
	111	50.4	5.1	176099	9	AC108201
	112	50.4	5.1	188823	2	AC016990
	113	50.4	5.1	201105	2	AC025668
	114	50.2	5.1	103202	2	AC087107
	115	50.2	5.1	112765	2	AC012129
c	116	50.2	5.1	137947	9	HS45I4
c	117	50.2	5.1	156394	9	AC105756
c	118	50.2	5.1	163277	2	AC084687
c	119	49.6	5.0	556	12	AF247170
c	120	49.6	5.0	44564	9	AL355871
	121	49.6	5.0	71653	9	HS133H11
c	122	49.6	5.0	78474	9	AC018762
c	123	49.6	5.0	101716	9	AC026807
c	124	49.6	5.0	151482	9	AC068765
	125	49.6	5.0	174028	2	AC067831
	126	49.6	5.0	181904	2	AC027623
	127	49.6	5.0	197828	2	AC034169
c	128	49.6	5.0	329643	2	AL512381
c	129	49.6	5.0	339168	2	AC078987
	130	49.4	5.0	91912	9	AL589935
	131	49.4	5.0	119452	9	AC096552
c	132	49.4	5.0	125020	9	AF429315
c	133	49.4	5.0	182082	2	AC092842
	134	49.2	5.0	125066	9	AC074286
c	135	49.2	5.0	173315	9	AL138957
	136	49.2	5.0	190221	9	AC025741
	137	49	4.9	153304	9	AC084834
c	138	49	4.9	160766	9	AC022679

AC117371 Homo sapi
AC073620 Homo sapi
AC011389 Homo sapi
AC073892 Homo sapi
AL138835 Human DNA
AC106871 Homo sapi
AC021271 Homo sapi
AC024482 Homo sapi
AC027460 Homo sapi
AC078965 Homo sapi
AC092380 Homo sapi
AC007511 Homo sapi
AC092371 Homo sapi
AC092136 Homo sapi
AC092125 Homo sapi
AC020743 Homo sapi
AL137879 Human DNA
AL606535 Human DNA
AC067733 Homo sapi
AC016706 Homo sapi
AC067734 Homo sapi
AJ009616 Homo sapi
AC005971 Homo sapi
AC073414 Homo sapi
AP001032 Homo sapi
AL359641 Human DNA
AC090857 Homo sapi
AC019137 Homo sapi
AP002757 Homo sapi
AL136140 Human DNA
AC112255 Homo sapi
AL359915 Human DNA
AC007312 Homo sapi
AB020873 Homo sapi
AL163011 Human chr
AL731897 Human DNA
AP006210 Homo sapi
AL691420 Human DNA
AC016204 Homo sapi
AC011767 Homo sapi
AL353649 Human DNA
AC009900 Homo sapi
AL354877 Human DNA
AC114971 Homo sapi
AC010780 Homo sapi
AC108201 Homo sapi
AC016990 Homo sapi
AC025668 Mus muscu
AC087107 Homo sapi
AC012129 Homo sapi
AL023581 Human DNA
AC105756 Homo sapi
AC084687 Homo sapi
AF247170 Synthetic
AL355871 Human DNA
AL024506 Human DNA
AC018762 Homo sapi
AC026807 Homo sapi
AC068765 Homo sapi
AC067831 Homo sapi
AC027623 Homo sapi
AC034169 Homo sapi
AL512381 Homo sapi
AC078987 Homo sapi
AL589935 Human DNA
AC096552 Homo sapi
AF429315 Homo sapi
AC092842 Homo sapi
AC074286 Homo sapi
AL138957 Human DNA
AC025741 Homo sapi
AC084834 Homo sapi
AC022679 Homo sapi

c	139	49	4.9	163260	2	AC091049
	140	49	4.9	186533	9	AC024606
	141	49	4.9	207203	2	AL512273
c	142	48.8	4.9	41255	9	HS439A6
c	143	48.8	4.9	75108	9	AC005363
c	144	48.8	4.9	81579	9	AE006640
c	145	48.8	4.9	92019	9	AL450320
c	146	48.8	4.9	124250	9	AC093387
c	147	48.8	4.9	153190	9	AC068597
	148	48.8	4.9	165990	2	AC073509
c	149	48.8	4.9	181570	2	AC068738
	150	48.6	4.9	2000	6	AX655393

AC091049 Homo sapi
AC024606 Homo sapi
AL512273 Homo sapi
AL031723 Human DNA
AC005363 Homo sapi
AE006640 Homo sapi
AL450320 Human DNA
AC093387 Homo sapi
AC068597 Homo sapi
AC073509 Homo sapi
AC068738 Homo sapi
AX655393 Sequence

ALIGNMENTS

RESULT 1
AR252529
LOCUS AR252529 992 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 222 from patent US 6478825.
ACCESSION AR252529
VERSION AR252529.1 GI:27300437
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 992)
AUTHORS Winterbottom,J.M., Shimp,L., Boyce,T.M. and Kaes,D.
TITLE Implant, method of making same and use of the implant for the treatment of bone defects
JOURNAL Patent: US 6478825-A 222 12-NOV-2002;
FEATURES Location/Qualifiers
source 1. .992
/organism="unknown"
BASE COUNT 235 a 301 c 282 g 174 t
ORIGIN

Query Match 100.0%; Score 992; DB 6; Length 992;
Best Local Similarity 100.0%; Pred. No. 1.8e-228;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
```

```
Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 2
AX358896
LOCUS AX358896 992 bp DNA linear PAT 13-FEB-2002
DEFINITION Sequence 149 from Patent WO0193983.
ACCESSION AX358896
VERSION AX358896.1 GI:18675349
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Baker,K.P., Desnoyers,L., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,J.C., Gurney,A.L., Smith,V., Stephan,J.P.,
Watanabe,C.K. and Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0193983-A 149 13-DEC-2001;
Genentech Inc. (US)
FEATURES Location/Qualifiers
source 1. .992
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 235 a 301 c 282 g 174 t
ORIGIN

Query Match 100.0%; Score 992; DB 6; Length 992;
Best Local Similarity 100.0%; Pred. No. 1.8e-228;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 3
AX362389
LOCUS AX362389 992 bp DNA linear PAT 15-FEB-2002
DEFINITION Sequence 149 from Patent WO0208288.
ACCESSION AX362389
VERSION AX362389.1 GI:18694652
KEYWORDS .
SOURCE Homo sapiens (human)

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Baker,K.P., Desnoyers,L., Gerritsen,M.E., Goddard,A.,
            Godowski,P.J., Grimaldi,J.C., Gurney,A.L., Smith,V., Stephan,J.P.,
            Watanabe,C.K. and Wood,W.I.
  TITLE     Secreted and transmembrane polypeptides and nucleic acids encoding
            the same
  JOURNAL   Patent: WO 0208288-A 149 31-JAN-2002;
            Genentech, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .992
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      235 a    301 c    282 g    174 t
ORIGIN

  Query Match           100.0%;  Score 992;  DB 6;  Length 992;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-228;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 4
AX403335
LOCUS       AX403335                992 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 222 from Patent WO0073454.
ACCESSION   AX403335
VERSION     AX403335.1  GI:21436883
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ashkenazi,A.J., Baker,K.P., Botstein,D., Desnoyers,L., Eaton,D.,
            Ferrara,N., Gerber,H., Gerritsen,M., Goddard,A., Godowski,P.,
            Grimaldi,C.J., Gurney,A.L., Kljavin,I., Napier,M.A., Pan,J.,
            Paoni,N.F., Roy,M., Stewart,T.A., Tumas,D., Watanabe,C.K.,
            Williams,P., Wood,W.I. and Zhang,Z.
  TITLE     Secreted and transmembrane polypeptides and nucleic acids encoding
            the same
  JOURNAL   Patent: WO 0073454-A 222 07-DEC-2000;
            Genentech Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .992
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      235 a    301 c    282 g    174 t
ORIGIN

  Query Match           100.0%;  Score 992;  DB 6;  Length 992;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-228;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
```

```
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
```

```
RESULT 5
BD107970
LOCUS       BD107970                505 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  EST and encoded human protein.
ACCESSION   BD107970
VERSION     BD107970.1  GI:23202788
KEYWORDS    JP 2002010789-A/47.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.
  TITLE     EST and encoded human protein
  JOURNAL   Patent: JP 2002010789-A 47 15-JAN-2002;
            GENSET CORP
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002010789-A/47
            PD   15-JAN-2002
            PF   07-AUG-2000 JP 2000280989
            PR   05-AUG-1999 US    60/147499
            PI   JEAN BAPUTIST DUMAS MILNE EDWARDS,SEVELIN JOBERT,JEAN EVE  PI
             GIORDANO
            PC   C12N15/09,C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19, PC
             C12N1/21,
            PC   C12N5/10,C12P21/02,C12P21/08,C12Q1/68,C12N15/00,C12N5/00, PC
            C12N15/00
            CC   Von Heijne matrix
            CC   score 8.89999961853027
            CC   seq LFCLAVLAASSFS/KA
            FH   Key             Location/Qualifiers
            FT   CDS             27. .503
            FT   sig_peptide     27. .80.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      106 a    170 c    136 g     92 t      1 others
ORIGIN
```

```
  Query Match           50.9%;  Score 504.6;  DB 6;  Length 505;
  Best Local Similarity  99.8%;  Pred. No. 6.5e-111;
  Matches  504;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         37 AGCAGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGC 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCAGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGC 60

Qy         97 TGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTG 120

Qy        157 CCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCAC 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCAC 180

Qy        217 CCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCA 240

Qy        277 AGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCA 300

Qy        337 GTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACA 396
              ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        301 GTCCAGACCTGCTCACCTACTTCTGCYGGGCGTCCTCCACCTCAGGTGCCCATGTGGACA 360

Qy        397 GTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCA 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCA 420

Qy        457 ACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCT 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCT 480

Qy        517 CGGGCAGCCCACCTATCACCAACAG 541
              |||||||||||||||||||||||||
Db        481 CGGGCAGCCCACCTATCACCAACAG 505
```

```
RESULT 6
AX588820
LOCUS       AX588820                502 bp    DNA     linear   PAT 24-JAN-2003
DEFINITION  Sequence 695 from Patent WO02083898.
ACCESSION   AX588820
VERSION     AX588820.1  GI:27900608
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Bejanin,S., Tanaka,H., Dumas Milne Edwards,J.B., Jobert,S. and
            Giordano,J.Y.
  TITLE     Full-length human cdnas encoding potentially secreted proteins
  JOURNAL   Patent: WO 02083898-A 695 24-OCT-2002;
```

```
           GENSET (FR)
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      105 a    169 c    135 g     92 t      1 others
ORIGIN
```

Query Match 50.6%; Score 501.6; DB 6; Length 502;
Best Local Similarity 99.8%; Pred. No. 3.4e-110;
Matches 501; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         40 AGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGG 99
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGG 60

Qy        100 CTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCT 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCT 120

Qy        160 ACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCC 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCC 180

Qy        220 AGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGA 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGA 240

Qy        280 AGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTC 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTC 300

Qy        340 CAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTG 399
              ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        301 CAGACCTGCTCACCTACTTCTGCYGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTG 360

Qy        400 CCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACT 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACT 420

Qy        460 TCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGG 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGG 480

Qy        520 GCAGCCCACCTATCACCAACAG 541
              |||||||||||||||||||||
Db        481 GCAGCCCACCTATCACCAACAG 502
```

```
RESULT 7
G60502/c
LOCUS       G60502                   414 bp    DNA     linear   STS 26-OCT-1999
DEFINITION  SHGC-33533 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G60502
VERSION     G60502.1  GI:6125821
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   Myers,R.M.
  TITLE     Human STSs (1999)
  JOURNAL   Unpublished (1999)
COMMENT
            Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu

            Primer A: GTGGCAAGAGCAGCTAAAAC
            Primer B: GGAATGGGGAGGTCAGAGG
            STS size: 150
            PCR Profile:
                    Initial incubation: 95 degrees C for 10 minutes
                    Denaturation:       94 degrees C for 30 seconds
                    Annealing:          60 degrees C for 30 seconds
                    Polymerization:     72 degrees C for 23 seconds
                    PCR Cycles:         30
                    Thermal Cycler:     Perkin Elmer 9700
            Protocol:
                    Template:                   25 ng
                    Primer:                     each 1 uM
                    dNTPs:                      each 200 uM
                    AmpliTaq Gold Polymerase:   0.07 units/ul
                    Total Vol:                  5 ul

            Buffer:
                    MgCl2:          2.5 mM
                    KCl:            50 mM
                    Tris-HCl:       10 mM
                    pH:             8.3

            Prepared with primer pairs provided by Sandoz, derived from H74303
            -- Washington University/Merck EST sequence.
FEATURES             Location/Qualifiers
     source          1. .414
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="17"
                     /clone_lib="Human"
     STS             1. .150
     primer_bind     1. .20
     primer_bind     complement(132. .150)
BASE COUNT       75 a    127 c    116 g     95 t      1 others
ORIGIN
```

Query Match 41.3%; Score 409.8; DB 11; Length 414;
Best Local Similarity 99.3%; Pred. No. 4.7e-88;
Matches 411; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy        556 ATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCC 615
              |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db        414 ATGGGCAGGTCCACCTGCAGCAGAGACCNATCCACAGGCAGCCTGCCAACTTCTCCTTCC 355

Qy        616 TGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGC 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 TGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGC 295

Qy        676 ACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTC 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 ACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTC 235

Qy        736 CCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAG 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 CCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAG 175

Qy        796 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        174 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 115

Qy        856 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCA 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCA 55

Qy        916 CTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 CTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1
```

```
RESULT 8
AC032035/c
```

```
LOCUS       AC032035              154840 bp    DNA     linear   HTG 08-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-141D15 map 17, *** SEQUENCING
            IN PROGRESS ***, 18 unordered pieces.
ACCESSION   AC032035
VERSION     AC032035.3  GI:9994161
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154840)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-141D15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154840)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 8, 2000 this sequence version replaced gi:7705196.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9138
                Center clone name: 141_D_15
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2373: contig of 2373 bp in length
            *     2374     2473: gap of 100 bp
            *     2474     4930: contig of 2457 bp in length
            *     4931     5030: gap of 100 bp
            *     5031     7215: contig of 2185 bp in length
            *     7216     7315: gap of 100 bp
            *     7316     9541: contig of 2226 bp in length
            *     9542     9641: gap of 100 bp
            *     9642    12211: contig of 2570 bp in length
            *    12212    12311: gap of 100 bp
            *    12312    14923: contig of 2612 bp in length
            *    14924    15023: gap of 100 bp
            *    15024    19094: contig of 4071 bp in length
            *    19095    19194: gap of 100 bp
            *    19195    22286: contig of 3092 bp in length
            *    22287    22386: gap of 100 bp
            *    22387    26147: contig of 3761 bp in length
            *    26148    26247: gap of 100 bp
            *    26248    31933: contig of 5686 bp in length
            *    31934    32033: gap of 100 bp
            *    32034    37339: contig of 5306 bp in length
            *    37340    37439: gap of 100 bp
            *    37440    42858: contig of 5419 bp in length
            *    42859    42958: gap of 100 bp
            *    42959    48810: contig of 5852 bp in length
            *    48811    48910: gap of 100 bp
            *    48911    56029: contig of 7119 bp in length
            *    56030    56129: gap of 100 bp
            *    56130    62189: contig of 6060 bp in length
            *    62190    62289: gap of 100 bp
            *    62290    82959: contig of 20670 bp in length
            *    82960    83059: gap of 100 bp
            *    83060   107059: contig of 24000 bp in length
            *   107060   107159: gap of 100 bp
            *   107160   154840: contig of 47681 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154840
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-141D15"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    36180 a  41741 c  41557 g  32874 t   2488 others
ORIGIN
```

```
  Query Match           30.3%;  Score 300.4;  DB 2;  Length 154840;
  Best Local Similarity   99.7%;  Pred. No. 1e-61;
  Matches  301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        131 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 1842

Qy        191 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 1782

Qy        251 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 1722

Qy        311 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTC 370
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       1721 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCTGGGCGTC 1662

Qy        371 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 1602

Qy        431 CA 432
              ||
Db       1601 CA 1600
```

```
RESULT 9
AC021593/c
LOCUS       AC021593              177738 bp    DNA     linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-153A23, complete sequence.
ACCESSION   AC021593
VERSION     AC021593.15  GI:22038664
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177738)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-153A23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177738)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 177738)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 177738)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 1, 2002 this sequence version replaced gi:21702950.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5781
                Center clone name: 153_A_23
            --------------.
FEATURES             Location/Qualifiers
     source          1. .177738
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-153A23"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(345. .630)
                     /rpt_family="AluJo"
     repeat_region   781. .1085
                     /rpt_family="AluSx"
     repeat_region   1731. .1953
                     /rpt_family="MER20"
     repeat_region   complement(2886. .3192)
                     /rpt_family="AluJo"
     repeat_region   3270. .3335
                     /rpt_family="ORSL"
     repeat_region   3476. .3751
                     /rpt_family="AluSx"
     repeat_region   complement(4214. .4517)
                     /rpt_family="AluSp"
     repeat_region   4554. .4842
                     /rpt_family="AluJb"
     repeat_region   complement(5056. .5343)
                     /rpt_family="AluSx"
     repeat_region   complement(5344. .5641)
                     /rpt_family="AluSx"
     repeat_region   complement(6335. .6523)
                     /rpt_family="MLT1J2"
     repeat_region   complement(6524. .6821)
                     /rpt_family="AluSx"
     repeat_region   complement(6822. .6894)
                     /rpt_family="MLT1J2"
     repeat_region   7748. .7970
                     /rpt_family="AluSp"
     repeat_region   8439. .8735
                     /rpt_family="AluSx"
     repeat_region   8736. .9033
                     /rpt_family="AluY"
     repeat_region   complement(9074. .9373)
                     /rpt_family="MER46B"
     repeat_region   9422. .9487
                     /rpt_family="AluS"
     repeat_region   9696. .9854
                     /rpt_family="MER20"
     unsure          9753. .9761
                     /note="<30 qual SNGL region"
     repeat_region   complement(10581. .10677)
                     /rpt_family="MIR"
     repeat_region   complement(12611. .12907)
                     /rpt_family="AluSx"
     unsure          14906. .14914
                     /note="PCR product sequence only"
     repeat_region   complement(15909. .16192)
                     /rpt_family="AluSq"
     repeat_region   17759. .17801
                     /rpt_family="GC_rich"
     repeat_region   17995. .18021
                     /rpt_family="(CCG)n"
     repeat_region   18569. .18880
                     /rpt_family="AluSp"
```

```
     repeat_region   18891. .19237
                     /rpt_family="AluSx"
     repeat_region   19239. .19387
                     /rpt_family="AluY"
     repeat_region   19388. .19675
                     /rpt_family="AluSc"
     repeat_region   19676. .19690
                     /rpt_family="AluY"
     repeat_region   complement(20303. .20438)
                     /rpt_family="L1MA9"
     repeat_region   complement(20799. .21088)
                     /rpt_family="AluSp"
     repeat_region   complement(21177. .21230)
                     /rpt_family="MIR"
     repeat_region   complement(22718. .23024)
                     /rpt_family="AluJb"
     repeat_region   complement(23053. .23147)
                     /rpt_family="MER53"
     repeat_region   complement(23349. .23524)
                     /rpt_family="AluJo"
     repeat_region   23525. .23820
                     /rpt_family="AluSx"
     repeat_region   complement(23821. .23936)
                     /rpt_family="AluJo"
     repeat_region   complement(23947. .24088)
                     /rpt_family="MIR"
     repeat_region   24216. .24315
                     /rpt_family="MIR"
     repeat_region   complement(24316. .24615)
                     /rpt_family="AluSx"
     repeat_region   24616. .24736
                     /rpt_family="MIR"
     repeat_region   complement(24742. .25021)
                     /rpt_family="AluSx"
     repeat_region   25728. .26024
                     /rpt_family="AluSx"
     repeat_region   complement(26044. .26353)
                     /rpt_family="AluSg"
     repeat_region   26354. .26416
                     /rpt_family="L1PA16"
     repeat_region   26458. .27006
                     /rpt_family="L1MB5"
     repeat_region   complement(27038. .27252)
                     /rpt_family="AluSq"
     repeat_region   complement(27256. .27450)
                     /rpt_family="AluSx"

  Query Match          30.3%;  Score 300.4;  DB 9;  Length 177738;
  Best Local Similarity   99.7%;  Pred. No. 1e-61;
  Matches  301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       131 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25656 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 25597

Qy       191 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25596 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 25537

Qy       251 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 310
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25536 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 25477

Qy       311 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTC 370
            ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     25476 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCTGGGCGTC 25417

Qy       371 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 430
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25416 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 25357

Qy       431 CA 432
             ||
```

```
Db     25356 CA 25355


RESULT 10
AC087645/c
LOCUS       AC087645              209751 bp    DNA     linear   PRI 05-FEB-2003
DEFINITION  Homo sapiens chromosome 17, clone RP11-219G17, complete sequence.
ACCESSION   AC087645
VERSION     AC087645.19  GI:27311503
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209751)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-219G17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209751)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 209751)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 209751)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
```

```
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 20, 2002 this sequence version replaced gi:26190573.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11992
                Center clone name: 219_G_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .209751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-219G17"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   95. .232
                     /rpt_family="C-rich"
     repeat_region   complement(634. .972)
                     /rpt_family="AluSx"
     repeat_region   complement(1653. .1760)
                     /rpt_family="MER5A"
     repeat_region   complement(1807. .1886)
                     /rpt_family="MIR"
     repeat_region   complement(2001. .2312)
                     /rpt_family="AluJo"
     repeat_region   2336. .2636
                     /rpt_family="AluSx"
     repeat_region   complement(3684. .3706)
                     /rpt_family="AluJo"
     repeat_region   3707. .3993
                     /rpt_family="AluSx"
     repeat_region   complement(3994. .4267)
                     /rpt_family="AluJo"
     repeat_region   5194. .5371
                     /rpt_family="AluSg/x"
     repeat_region   5606. .5738
                     /rpt_family="FRAM"
     repeat_region   complement(5739. .5850)
                     /rpt_family="Alu"
     repeat_region   complement(5865. .5999)
                     /rpt_family="AluSg/x"
     repeat_region   6293. .6600
                     /rpt_family="AluSp"
     repeat_region   complement(7562. .7879)
                     /rpt_family="AluY"
     repeat_region   8178. .8326
                     /rpt_family="L1ME4a"
     repeat_region   8327. .8617
                     /rpt_family="AluSx"
     repeat_region   8618. .8721
                     /rpt_family="L1ME4a"
     repeat_region   8799. .9231
                     /rpt_family="L3"
     repeat_region   9248. .9488
                     /rpt_family="L3"
     repeat_region   9996. .10097
                     /rpt_family="AluSg/x"
     repeat_region   complement(10100. .10397)
                     /rpt_family="AluY"
     unsure          10159. .10164
                     /note="<30 qual SNGL region"
     repeat_region   10399. .10579
                     /rpt_family="AluSg/x"
     repeat_region   10580. .10908
                     /rpt_family="AluSx"
     repeat_region   10909. .10919
                     /rpt_family="AluSg/x"
     repeat_region   11385. .11522
                     /rpt_family="MIR"
     repeat_region   complement(11696. .11997)
                     /rpt_family="AluY"
     repeat_region   complement(12826. .12967)
                     /rpt_family="L1MC4"
     repeat_region   complement(12968. .13262)
                     /rpt_family="AluJb"
     repeat_region   complement(13263. .13344)
                     /rpt_family="L1MC4"
     repeat_region   complement(13354. .13532)
                     /rpt_family="AluJb"
     repeat_region   complement(13533. .13820)
                     /rpt_family="AluSx"
     repeat_region   complement(13821. .13890)
                     /rpt_family="L1MC4"
     repeat_region   14283. .14348
                     /rpt_family="MER45C"
     repeat_region   complement(14349. .14525)
                     /rpt_family="AluSg"
     repeat_region   complement(14526. .14828)
                     /rpt_family="AluY"
     repeat_region   complement(14829. .14962)
                     /rpt_family="AluSg"
     repeat_region   14963. .14999
                     /rpt_family="MER45C"
     repeat_region   15000. .15299
                     /rpt_family="AluSx"
     repeat_region   15303. .15598
                     /rpt_family="AluSx"
     repeat_region   15599. .15752
                     /rpt_family="MER45C"
     repeat_region   complement(16198. .16332)
                     /rpt_family="MIR"
     repeat_region   16832. .16961
                     /rpt_family="MIR"
     unsure          17736. .17810
                     /note="single clone coverage"
     repeat_region   17794. .17975
                     /rpt_family="CT-rich"
     repeat_region   18235. .18553
                     /rpt_family="AluSq"

  Query Match             30.3%;  Score 300.4;  DB 9;  Length 209751;
  Best Local Similarity   99.7%;  Pred. No. 1e-61;
  Matches  301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        131 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181172 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 181113

Qy        191 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181112 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 181053

Qy        251 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 310
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181052 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 180993

Qy         311 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTC 370
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      180992 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCTGGGCGTC 180933

Qy         371 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 430
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      180932 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 180873

Qy         431 CA 432
                ||
Db      180872 CA 180871


RESULT 11
AC010532
LOCUS       AC010532               229426 bp    DNA     linear   HTG 05-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-219G17, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC010532
VERSION     AC010532.2  GI:6758797
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229426)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229426)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 26, 2000 this sequence version replaced gi:5882406.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            --------.
            * NOTE: This record contains 99 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      563: contig of 563 bp in length
            *                    gap of unknown length
            *      564     1436: contig of 873 bp in length
            *                    gap of unknown length
            *     1437     1649: contig of 213 bp in length
            *                    gap of unknown length
            *     1650     2523: contig of 874 bp in length
            *                    gap of unknown length
            *     2524     3286: contig of 763 bp in length
            *                    gap of unknown length
            *     3287     3613: contig of 327 bp in length
            *                    gap of unknown length
            *     3614     5430: contig of 1817 bp in length
            *                    gap of unknown length
            *     5431     6681: contig of 1251 bp in length
            *                    gap of unknown length
            *     6682     7357: contig of 676 bp in length
            *                    gap of unknown length
            *     7358     8117: contig of 760 bp in length
            *                    gap of unknown length
            *     8118     8718: contig of 601 bp in length
            *                    gap of unknown length
            *     8719     9464: contig of 746 bp in length
            *                    gap of unknown length
            *     9465     9964: contig of 500 bp in length
            *                    gap of unknown length
            *     9965    10959: contig of 995 bp in length
            *                    gap of unknown length
            *    10960    11873: contig of 914 bp in length
            *                    gap of unknown length
            *    11874    12656: contig of 783 bp in length
            *                    gap of unknown length
            *    12657    13674: contig of 1018 bp in length
            *                    gap of unknown length
            *    13675    14935: contig of 1261 bp in length
            *                    gap of unknown length
            *    14936    16423: contig of 1488 bp in length
            *                    gap of unknown length
            *    16424    17186: contig of 763 bp in length
            *                    gap of unknown length
            *    17187    18079: contig of 893 bp in length
            *                    gap of unknown length
            *    18080    18813: contig of 734 bp in length
            *                    gap of unknown length
            *    18814    19729: contig of 916 bp in length
            *                    gap of unknown length
            *    19730    20675: contig of 946 bp in length
            *                    gap of unknown length
            *    20676    21721: contig of 1046 bp in length
            *                    gap of unknown length
            *    21722    23111: contig of 1390 bp in length
            *                    gap of unknown length
            *    23112    25188: contig of 2077 bp in length
            *                    gap of unknown length
            *    25189    26588: contig of 1400 bp in length
            *                    gap of unknown length
            *    26589    27644: contig of 1056 bp in length
            *                    gap of unknown length
            *    27645    29372: contig of 1728 bp in length
            *                    gap of unknown length
            *    29373    29495: contig of 123 bp in length
            *                    gap of unknown length
            *    29496    30431: contig of 936 bp in length
            *                    gap of unknown length
            *    30432    31449: contig of 1018 bp in length
            *                    gap of unknown length
            *    31450    32983: contig of 1534 bp in length
            *                    gap of unknown length
            *    32984    33162: contig of 179 bp in length
            *                    gap of unknown length
            *    33163    33975: contig of 813 bp in length
            *                    gap of unknown length
            *    33976    35031: contig of 1056 bp in length
            *                    gap of unknown length
            *    35032    35709: contig of 678 bp in length
            *                    gap of unknown length
            *    35710    37308: contig of 1599 bp in length
            *                    gap of unknown length
            *    37309    38107: contig of 799 bp in length
            *                    gap of unknown length
            *    38108    39102: contig of 995 bp in length
            *                    gap of unknown length
            *    39103    39591: contig of 489 bp in length
            *                    gap of unknown length
            *    39592    41151: contig of 1560 bp in length
            *                    gap of unknown length
            *    41152    42842: contig of 1691 bp in length
            *                    gap of unknown length
            *    42843    43828: contig of 986 bp in length
            *                    gap of unknown length
            *    43829    44009: contig of 181 bp in length
            *                    gap of unknown length
```

```
*    44010    44874: contig of 865 bp in length
*                    gap of unknown length
*    44875    45191: contig of 317 bp in length
*                    gap of unknown length
*    45192    46147: contig of 956 bp in length
*                    gap of unknown length
*    46148    47504: contig of 1357 bp in length
*                    gap of unknown length
*    47505    49091: contig of 1587 bp in length
*                    gap of unknown length
*    49092    50259: contig of 1168 bp in length
*                    gap of unknown length
*    50260    51361: contig of 1102 bp in length
*                    gap of unknown length
*    51362    52631: contig of 1270 bp in length
*                    gap of unknown length
*    52632    53607: contig of 976 bp in length
*                    gap of unknown length
*    53608    54820: contig of 1213 bp in length
*                    gap of unknown length
*    54821    56600: contig of 1780 bp in length
*                    gap of unknown length
*    56601    58236: contig of 1636 bp in length
*                    gap of unknown length
*    58237    59716: contig of 1480 bp in length
*                    gap of unknown length
*    59717    62221: contig of 2505 bp in length
*                    gap of unknown length
*    62222    65085: contig of 2864 bp in length
*                    gap of unknown length
*    65086    66739: contig of 1654 bp in length
*                    gap of unknown length
*    66740    68359: contig of 1620 bp in length
*                    gap of unknown length
*    68360    70014: contig of 1655 bp in length
*                    gap of unknown length
*    70015    71808: contig of 1794 bp in length
*                    gap of unknown length
*    71809    73581: contig of 1773 bp in length
*                    gap of unknown length
*    73582    75574: contig of 1993 bp in length
*                    gap of unknown length
*    75575    78468: contig of 2894 bp in length
*                    gap of unknown length
*    78469    79231: contig of 763 bp in length
*                    gap of unknown length
*    79232    83240: contig of 4009 bp in length
*                    gap of unknown length
*    83241    86157: contig of 2917 bp in length
*                    gap of unknown length
*    86158    89150: contig of 2993 bp in length
*                    gap of unknown length
*    89151    92671: contig of 3521 bp in length
*                    gap of unknown length
*    92672    95442: contig of 2771 bp in length
*                    gap of unknown length
*    95443    98476: contig of 3034 bp in length
*                    gap of unknown length
*    98477   100815: contig of 2339 bp in length
*                    gap of unknown length
*   100816   103005: contig of 2190 bp in length
*                    gap of unknown length
*   103006   105514: contig of 2509 bp in length
*                    gap of unknown length
*   105515   108704: contig of 3190 bp in length
*                    gap of unknown length
*   108705   111097: contig of 2393 bp in length
*                    gap of unknown length
*   111098   114888: contig of 3791 bp in length
*                    gap of unknown length
*   114889   117423: contig of 2535 bp in length
*                    gap of unknown length
```

```
  Query Match             29.5%;  Score 292.8;  DB 2;  Length 229426;
  Best Local Similarity   97.7%;  Pred. No. 6.8e-60;
  Matches  297;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy        688 CAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 747
              || || |     | ||||||||||||||||||||||||||||||||||||||||||||||
Db     135296 CACTGCTTTGTCAAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 135355

Qy        748 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135356 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 135415

Qy        808 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135416 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 135475

Qy        868 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135476 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 135535

Qy        928 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAA 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135536 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAA 135595

Qy        988 AAAA 991
              ||||
Db     135596 AAAA 135599
```

```
RESULT 12
AC032035
LOCUS       AC032035              154840 bp    DNA     linear   HTG 08-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-141D15 map 17, *** SEQUENCING
            IN PROGRESS ***, 18 unordered pieces.
ACCESSION   AC032035
VERSION     AC032035.3  GI:9994161
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154840)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-141D15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154840)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 8, 2000 this sequence version replaced gi:7705196.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ---------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          ---------- Project Information
              Center project name: L9138
              Center clone name: 141_D_15
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 18 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     2373: contig of 2373 bp in length
          *     2374     2473: gap of 100 bp
          *     2474     4930: contig of 2457 bp in length
          *     4931     5030: gap of 100 bp
          *     5031     7215: contig of 2185 bp in length
          *     7216     7315: gap of 100 bp
          *     7316     9541: contig of 2226 bp in length
          *     9542     9641: gap of 100 bp
          *     9642    12211: contig of 2570 bp in length
          *    12212    12311: gap of 100 bp
          *    12312    14923: contig of 2612 bp in length
          *    14924    15023: gap of 100 bp
          *    15024    19094: contig of 4071 bp in length
          *    19095    19194: gap of 100 bp
          *    19195    22286: contig of 3092 bp in length
          *    22287    22386: gap of 100 bp
          *    22387    26147: contig of 3761 bp in length
          *    26148    26247: gap of 100 bp
          *    26248    31933: contig of 5686 bp in length
          *    31934    32033: gap of 100 bp
          *    32034    37339: contig of 5306 bp in length
          *    37340    37439: gap of 100 bp
          *    37440    42858: contig of 5419 bp in length
          *    42859    42958: gap of 100 bp
          *    42959    48810: contig of 5852 bp in length
          *    48811    48910: gap of 100 bp
          *    48911    56029: contig of 7119 bp in length
          *    56030    56129: gap of 100 bp
          *    56130    62189: contig of 6060 bp in length
          *    62190    62289: gap of 100 bp
          *    62290    82959: contig of 20670 bp in length
          *    82960    83059: gap of 100 bp
          *    83060   107059: contig of 24000 bp in length
          *   107060   107159: gap of 100 bp
          *   107160   154840: contig of 47681 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154840
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-141D15"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    36180 a  41741 c  41557 g  32874 t   2488 others
ORIGIN

  Query Match             29.4%;  Score 291.2;  DB 2;  Length 154840;
  Best Local Similarity   97.4%;  Pred. No. 1.7e-59;
  Matches  296;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        688 CAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 747
              || || |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db      86135 CACTGCTTTGTCAAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 86194

Qy        748 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 807
              ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db      86195 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTTTGAGTGAAGAGGAGTTTGGGG 86254

Qy        808 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86255 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 86314

Qy        868 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86315 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 86374

Qy        928 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAA 987
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86375 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAA 86434

Qy        988 AAAA 991
              ||||
Db      86435 AAAA 86438


RESULT 13
AC010532/c
LOCUS       AC010532              229426 bp    DNA     linear   HTG 05-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-219G17, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC010532
VERSION     AC010532.2  GI:6758797
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 229426)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 229426)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 26, 2000 this sequence version replaced gi:5882406.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            --------.
            * NOTE: This record contains 99 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      563: contig of 563 bp in length
            *                    gap of unknown length
            *      564     1436: contig of 873 bp in length
            *                    gap of unknown length
            *     1437     1649: contig of 213 bp in length
            *                    gap of unknown length
            *     1650     2523: contig of 874 bp in length
            *                    gap of unknown length
            *     2524     3286: contig of 763 bp in length
            *                    gap of unknown length
            *     3287     3613: contig of 327 bp in length
            *                    gap of unknown length
            *     3614     5430: contig of 1817 bp in length
            *                    gap of unknown length
```

```
*     5431     6681: contig of 1251 bp in length
*                    gap of unknown length
*     6682     7357: contig of 676 bp in length
*                    gap of unknown length
*     7358     8117: contig of 760 bp in length
*                    gap of unknown length
*     8118     8718: contig of 601 bp in length
*                    gap of unknown length
*     8719     9464: contig of 746 bp in length
*                    gap of unknown length
*     9465     9964: contig of 500 bp in length
*                    gap of unknown length
*     9965    10959: contig of 995 bp in length
*                    gap of unknown length
*    10960    11873: contig of 914 bp in length
*                    gap of unknown length
*    11874    12656: contig of 783 bp in length
*                    gap of unknown length
*    12657    13674: contig of 1018 bp in length
*                    gap of unknown length
*    13675    14935: contig of 1261 bp in length
*                    gap of unknown length
*    14936    16423: contig of 1488 bp in length
*                    gap of unknown length
*    16424    17186: contig of 763 bp in length
*                    gap of unknown length
*    17187    18079: contig of 893 bp in length
*                    gap of unknown length
*    18080    18813: contig of 734 bp in length
*                    gap of unknown length
*    18814    19729: contig of 916 bp in length
*                    gap of unknown length
*    19730    20675: contig of 946 bp in length
*                    gap of unknown length
*    20676    21721: contig of 1046 bp in length
*                    gap of unknown length
*    21722    23111: contig of 1390 bp in length
*                    gap of unknown length
*    23112    25188: contig of 2077 bp in length
*                    gap of unknown length
*    25189    26588: contig of 1400 bp in length
*                    gap of unknown length
*    26589    27644: contig of 1056 bp in length
*                    gap of unknown length
*    27645    29372: contig of 1728 bp in length
*                    gap of unknown length
*    29373    29495: contig of 123 bp in length
*                    gap of unknown length
*    29496    30431: contig of 936 bp in length
*                    gap of unknown length
*    30432    31449: contig of 1018 bp in length
*                    gap of unknown length
*    31450    32983: contig of 1534 bp in length
*                    gap of unknown length
*    32984    33162: contig of 179 bp in length
*                    gap of unknown length
*    33163    33975: contig of 813 bp in length
*                    gap of unknown length
*    33976    35031: contig of 1056 bp in length
*                    gap of unknown length
*    35032    35709: contig of 678 bp in length
*                    gap of unknown length
*    35710    37308: contig of 1599 bp in length
*                    gap of unknown length
*    37309    38107: contig of 799 bp in length
*                    gap of unknown length
*    38108    39102: contig of 995 bp in length
*                    gap of unknown length
*    39103    39591: contig of 489 bp in length
*                    gap of unknown length
*    39592    41151: contig of 1560 bp in length
*                    gap of unknown length
*    41152    42842: contig of 1691 bp in length
*                    gap of unknown length
*    42843    43828: contig of 986 bp in length
*                    gap of unknown length
*    43829    44009: contig of 181 bp in length
*                    gap of unknown length
*    44010    44874: contig of 865 bp in length
*                    gap of unknown length
*    44875    45191: contig of 317 bp in length
*                    gap of unknown length
*    45192    46147: contig of 956 bp in length
*                    gap of unknown length
*    46148    47504: contig of 1357 bp in length
*                    gap of unknown length
*    47505    49091: contig of 1587 bp in length
*                    gap of unknown length
*    49092    50259: contig of 1168 bp in length
*                    gap of unknown length
*    50260    51361: contig of 1102 bp in length
*                    gap of unknown length
*    51362    52631: contig of 1270 bp in length
*                    gap of unknown length
*    52632    53607: contig of 976 bp in length
*                    gap of unknown length
*    53608    54820: contig of 1213 bp in length
*                    gap of unknown length
*    54821    56600: contig of 1780 bp in length
*                    gap of unknown length
*    56601    58236: contig of 1636 bp in length
*                    gap of unknown length
*    58237    59716: contig of 1480 bp in length
*                    gap of unknown length
*    59717    62221: contig of 2505 bp in length
*                    gap of unknown length
*    62222    65085: contig of 2864 bp in length
*                    gap of unknown length
*    65086    66739: contig of 1654 bp in length
*                    gap of unknown length
*    66740    68359: contig of 1620 bp in length
*                    gap of unknown length
*    68360    70014: contig of 1655 bp in length
*                    gap of unknown length
*    70015    71808: contig of 1794 bp in length
*                    gap of unknown length
*    71809    73581: contig of 1773 bp in length
*                    gap of unknown length
*    73582    75574: contig of 1993 bp in length
*                    gap of unknown length
*    75575    78468: contig of 2894 bp in length
*                    gap of unknown length
*    78469    79231: contig of 763 bp in length
*                    gap of unknown length
*    79232    83240: contig of 4009 bp in length
*                    gap of unknown length
*    83241    86157: contig of 2917 bp in length
*                    gap of unknown length
*    86158    89150: contig of 2993 bp in length
*                    gap of unknown length
*    89151    92671: contig of 3521 bp in length
*                    gap of unknown length
*    92672    95442: contig of 2771 bp in length
*                    gap of unknown length
*    95443    98476: contig of 3034 bp in length
*                    gap of unknown length
*    98477   100815: contig of 2339 bp in length
*                    gap of unknown length
*   100816   103005: contig of 2190 bp in length
*                    gap of unknown length
*   103006   105514: contig of 2509 bp in length
*                    gap of unknown length
*   105515   108704: contig of 3190 bp in length
*                    gap of unknown length
*   108705   111097: contig of 2393 bp in length
*                    gap of unknown length
```

```
            *   111098   114888: contig of 3791 bp in length
            *                   gap of unknown length
            *   114889   117423: contig of 2535 bp in length
            *                   gap of unknown length

  Query Match           18.1%;  Score 179.4;  DB 2;  Length 229426;
  Best Local Similarity  96.8%;  Pred. No. 1.5e-32;
  Matches  183;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       244 CCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGG 303
            |||| | | | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1331 CGCTCGGGGAACCCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGG 1272

Qy       304 CCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCC 363
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1271 CCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCT 1212

Qy       364 GGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGC 423
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1211 GGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGC 1152

Qy       424 TGTGGTCCA 432
            |||||||||
Db      1151 TGTGGTCCA 1143


RESULT 14
AL645856
LOCUS       AL645856               116300 bp    DNA     linear   ROD 24-AUG-2002
DEFINITION  Mouse DNA sequence from clone RP23-151N19 on chromosome 11,
            complete sequence.
ACCESSION   AL645856
VERSION     AL645856.5  GI:22531400
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 116300)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2002 this sequence version replaced gi:21738435.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-151N19 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.

FEATURES             Location/Qualifiers
     source          1. .116300
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-151N19"
                     /clone_lib="RPCI-23"
BASE COUNT    29288 a  28679 c  29788 g  28545 t
ORIGIN

  Query Match           16.5%;  Score 163.4;  DB 10;  Length 116300;
  Best Local Similarity  73.3%;  Pred. No. 1.1e-28;
  Matches  209;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       148 TCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCT 207
            || |||||||||||||||| |||||||||  ||| || ||||  |||||||  |||||||
Db     93425 TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCT 93484

Qy       208 GCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCA 267
            ||  ||| ||  || |  | | |||||||| || || |||   |  |  |   |||
Db     93485 GCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCC 93544

Qy       268 AGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACAC 327
            ||||||  || ||||| ||| |   |  || |||||||||||||||  ||||  ||||
Db     93545 TGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCA 93604

Qy       328 TCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCC 387
            |||||||||| |||||||||||||||||| ||||| |||  |||| | ||| ||  ||
Db     93605 TCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCT 93664

Qy       388 ATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCA 432
            |||   |||   |||||| |||||| ||  ||| |||||| | |
Db     93665 ATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTGTGGGCTA 93709


RESULT 15
AL954690
LOCUS       AL954690               197194 bp    DNA     linear   HTG 15-NOV-2002
DEFINITION  Mus musculus chromosome 11 clone RP23-386A4, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AL954690
VERSION     AL954690.2  GI:25136861
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 197194)
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 19, 2002 this sequence version replaced gi:25045706.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM386A4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 195990 bases at least Q40
            Consensus quality: 196234 bases at least Q30
            Consensus quality: 196338 bases at least Q20
            Insert size: 197194; sum-of-contigs
```

```
            Insert size: 189681; 3.6% error; agarose-fp
            Quality coverage: 7.24x in Q20 bases; sum-of-contigs Quality
            coverage: 7.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
     source           1. .197194
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10090"
                      /chromosome="11"
                      /clone="RP23-386A4"
                      /clone_lib="RPCI-23"
     misc_feature     1. .197194
                      /note="assembly_fragment:02398
                      clone_end:SP6
                      vector_side:left"
BASE COUNT     46441 a  51277 c  51550 g  47922 t      4 others
ORIGIN

  Query Match             16.5%;  Score 163.4;  DB 2;  Length 197194;
  Best Local Similarity   73.3%;  Pred. No. 1.1e-28;
  Matches  209;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy        148 TCTCCATTGCCTACAAAGTCCTGGAAGTTTCCCCAAAGGCCGCTGGGTGCTCATAACCT 207
              || ||||||||||||||| |||||||||  ||| || ||||  |||||| |||||||
Db      20120 TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCT 20179

Qy        208 GCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCA 267
              ||  ||| ||  || |  | | |||||||||| || || |||    |  | |       |||
Db      20180 GCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCC 20239

Qy        268 AGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACAC 327
              |||||| || ||||| ||| |    |  ||  |||||||||||||||| ||||  ||||
Db      20240 TGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCA 20299

Qy        328 TCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCC 387
              |||||||||| |||||||||||||||||| ||||| |||  |||| | ||| ||  ||
Db      20300 TCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCT 20359

Qy        388 ATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCA 432
              |||    |||   |||||| |||||| ||  ||| |||||| | |
Db      20360 ATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTGTGGGCTA 20404


RESULT 16
AC100565
LOCUS       AC100565               49532 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-154K21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC100565
VERSION     AC100565.1  GI:17047931
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 49532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-154K21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 49532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L15700
                Center clone name: 154_K_21
            --------------.
            * NOTE: This record contains 48 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      986: contig of 986 bp in length
            *      987     1086: gap of 100 bp
            *     1087     2009: contig of 923 bp in length
            *     2010     2109: gap of 100 bp
            *     2110     3007: contig of 898 bp in length
            *     3008     3107: gap of 100 bp
            *     3108     3866: contig of 759 bp in length
            *     3867     3966: gap of 100 bp
            *     3967     4874: contig of 908 bp in length
            *     4875     4974: gap of 100 bp
            *     4975     5839: contig of 865 bp in length
            *     5840     5939: gap of 100 bp
            *     5940     7017: contig of 1078 bp in length
            *     7018     7117: gap of 100 bp
            *     7118     8079: contig of 962 bp in length
            *     8080     8179: gap of 100 bp
            *     8180     9139: contig of 960 bp in length
            *     9140     9239: gap of 100 bp
            *     9240    10254: contig of 1015 bp in length
            *    10255    10354: gap of 100 bp
            *    10355    11289: contig of 935 bp in length
            *    11290    11389: gap of 100 bp
            *    11390    12334: contig of 945 bp in length
            *    12335    12434: gap of 100 bp
            *    12435    13410: contig of 976 bp in length
            *    13411    13510: gap of 100 bp
            *    13511    14445: contig of 935 bp in length
            *    14446    14545: gap of 100 bp
            *    14546    15272: contig of 727 bp in length
            *    15273    15372: gap of 100 bp
            *    15373    16394: contig of 1022 bp in length
            *    16395    16494: gap of 100 bp
            *    16495    17461: contig of 967 bp in length
            *    17462    17561: gap of 100 bp
```

```
            *    17562    18545: contig of 984 bp in length
            *    18546    18645: gap of 100 bp
            *    18646    19600: contig of 955 bp in length
            *    19601    19700: gap of 100 bp
            *    19701    20670: contig of 970 bp in length
            *    20671    20770: gap of 100 bp
            *    20771    21739: contig of 969 bp in length
            *    21740    21839: gap of 100 bp
            *    21840    22669: contig of 830 bp in length
            *    22670    22769: gap of 100 bp
            *    22770    23763: contig of 994 bp in length
            *    23764    23863: gap of 100 bp
            *    23864    24820: contig of 957 bp in length
            *    24821    24920: gap of 100 bp
            *    24921    25847: contig of 927 bp in length
            *    25848    25947: gap of 100 bp
            *    25948    26787: contig of 840 bp in length
            *    26788    26887: gap of 100 bp
            *    26888    27831: contig of 944 bp in length
            *    27832    27931: gap of 100 bp
            *    27932    28881: contig of 950 bp in length
            *    28882    28981: gap of 100 bp
            *    28982    29906: contig of 925 bp in length
            *    29907    30006: gap of 100 bp
            *    30007    30921: contig of 915 bp in length
            *    30922    31021: gap of 100 bp
            *    31022    31757: contig of 736 bp in length
            *    31758    31857: gap of 100 bp
            *    31858    32798: contig of 941 bp in length
            *    32799    32898: gap of 100 bp
            *    32899    33835: contig of 937 bp in length
            *    33836    33935: gap of 100 bp
            *    33936    34860: contig of 925 bp in length
            *    34861    34960: gap of 100 bp
            *    34961    35887: contig of 927 bp in length
            *    35888    35987: gap of 100 bp
            *    35988    36990: contig of 1003 bp in length
            *    36991    37090: gap of 100 bp
            *    37091    38085: contig of 995 bp in length
            *    38086    38185: gap of 100 bp
            *    38186    39172: contig of 987 bp in length
            *    39173    39272: gap of 100 bp
            *    39273    40215: contig of 943 bp in length
            *    40216    40315: gap of 100 bp
            *    40316    41267: contig of 952 bp in length
            *    41268    41367: gap of 100 bp
            *    41368    42271: contig of 904 bp in length
            *    42272    42371: gap of 100 bp
            *    42372    43302: contig of 931 bp in length
            *    43303    43402: gap of 100 bp
            *    43403    44283: contig of 881 bp in length
            *    44284    44383: gap of 100 bp
            *    44384    45395: contig of 1012 bp in length
            *    45396    45495: gap of 100 bp
            *    45496    46390: contig of 895 bp in length
            *    46391    46490: gap of 100 bp
            *    46491    47408: contig of 918 bp in length
            *    47409    47508: gap of 100 bp
            *    47509    48438: contig of 930 bp in length
            *    48439    48538: gap of 100 bp
            *    48539    49532: contig of 994 bp in length.
FEATURES             Location/Qualifiers
     source          1. .49532
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-154K21"
                     /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT    11202 a  11441 c  10738 g  11027 t   5124 others
ORIGIN

  Query Match           16.1%;  Score 160.2;  DB 2;  Length 49532;
  Best Local Similarity  72.6%;  Pred. No. 6.6e-28;
```

```
  Matches  207;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy       148 TCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCT 207
            ||  |||||||||||||| |||||||||  ||| || ||||  ||||||| |||||||
Db      9291 TCACTCTTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCT 9350

Qy       208 GCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCA 267
            ||  ||| ||  || |  | | |||||||| || || |||   |  | |      |||
Db      9351 GCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCC 9410

Qy       268 AGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACAC 327
            ||||||  || |||||  ||| |   |  ||  ||||||||||||||||  ||||  ||||
Db      9411 TGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCA 9470

Qy       328 TCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCC 387
            |||||||||  |||||||||||||||||||  |||||  |||  ||||  |  |||  ||  ||
Db      9471 TCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCT 9530

Qy       388 ATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCA 432
            |||    |||   |||||| ||||||  ||  ||| ||||||  | |
Db      9531 ATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTGTGGGCTA 9575


RESULT 17
G43545/c
LOCUS       G43545                   150 bp    mRNA    linear   STS 28-JAN-1999
DEFINITION  WIAF-2405-STS Human THudson EST Homo sapiens STS cDNA, sequence
            tagged site.
ACCESSION   G43545
VERSION     G43545.1  GI:4192462
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
            Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
            Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
            Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
            Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
            Lander,E.S.
  TITLE     Large-scale identification, mapping, and genotyping of
            single-nucleotide polymorphisms in the human genome
  JOURNAL   Science 280 (5366), 1077-1082 (1998)
  MEDLINE   98248615
   PUBMED   9582121
COMMENT     Synonyms: EST376246a, EST376246
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu
            Primer A: GTGGCAAGAGCAGCTAAAAC
            Primer B: GGAATGGGGAGGTCAGAGG
            STS size: 150
            PCR Profile:
                    Presoak:            94 degrees C for 4.00 minutes
                    Denaturation:       94 degrees C for 50.0 seconds
                    Annealing:          58 degrees C for 1.50 minutes
                    Polymerization:     72 degrees C for 1.00 minutes
                    PCR Cycles:         30
                    Thermal Cycler: custom built by IAS, Costar, Cambridge MA
            Protocol:
                  Template:       10 ng
                  Primer:         each 5 pM
                  dNTPs:          4 nM
                  Taq Polymerase: 0.5 U
                  Total Vol:      20 uL

            Buffer:
```

```
                Mg2+:          1.5 mM
                KCl:            50 mM
                Tris-HCl:       10 mM
                Gelatin:      .001 %.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human THudson EST"
                     /note="STSs derived from sequences in dbEST and the
                     Unigene collection."
     STS             1. .150
     primer_bind     1. .20
     primer_bind     complement(132. .150)
BASE COUNT       30 a     46 c     33 g     40 t      1 others
ORIGIN

  Query Match             15.0%;  Score 149;  DB 11;  Length 150;
  Best Local Similarity   99.3%;  Pred. No. 3.8e-25;
  Matches  149;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 91

Qy        880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         90 CAAGCACGGCAGAGGACTGCAGGCCATCAGCNTGCACTGTTCGTATTTGGAGTTCATGCA 31

Qy        940 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              ||||||||||||||||||||||||||||||
Db         30 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1


RESULT 18
G43546/c
LOCUS       G43546                   150 bp    mRNA    linear   STS 28-JAN-1999
DEFINITION  WIAF-2407-STS Human THudson EST Homo sapiens STS cDNA, sequence
            tagged site.
ACCESSION   G43546
VERSION     G43546.1  GI:4192463
KEYWORDS    STS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R.,
            Ghandour,G., Perkins,N., Winchester,E., Spencer,J., Kruglyak,L.,
            Stein,L., Hsie,L., Topaloglou,T., Hubbell,E., Robinson,E.,
            Mittmann,M., Morris,M.S., Shen,N., Kilburn,D., Rioux,J.,
            Nusbaum,C., Rozen,S., Hudson,T.J., Lipshutz,R., Chee,M. and
            Lander,E.S.
  TITLE     Large-scale identification, mapping, and genotyping of
            single-nucleotide polymorphisms in the human genome
  JOURNAL   Science 280 (5366), 1077-1082 (1998)
  MEDLINE   98248615
   PUBMED   9582121
COMMENT     Synonyms: EST376246b, EST376246
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu
            Primer A: GTGGCAAGAGCAGCTAAAAC
            Primer B: GGAATGGGGAGGTCAGAGG
            STS size: 150
            PCR Profile:
                    Presoak:            94 degrees C for 4.00 minutes
                    Denaturation:       94 degrees C for 50.0 seconds
                    Annealing:          58 degrees C for 1.50 minutes
                    Polymerization:     72 degrees C for 1.00 minutes
                    PCR Cycles:         30
                    Thermal Cycler: custom built by IAS, Costar, Cambridge MA
            Protocol:
                    Template:       10 ng
                    Primer:         each 5 pM
                    dNTPs:          4 nM
                    Taq Polymerase: 0.5 U
                    Total Vol:      20 uL

            Buffer:
                    Mg2+:          1.5 mM
                    KCl:            50 mM
                    Tris-HCl:       10 mM
                    Gelatin:      .001 %.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human THudson EST"
                     /note="STSs derived from sequences in dbEST and the
                     Unigene collection."
     STS             1. .150
     primer_bind     1. .20
     primer_bind     complement(132. .150)
BASE COUNT       30 a     47 c     32 g     40 t      1 others
ORIGIN

  Query Match             15.0%;  Score 149;  DB 11;  Length 150;
  Best Local Similarity   99.3%;  Pred. No. 3.8e-25;
  Matches  149;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        150 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAANCGTCCAGAGAGC 91

Qy        880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 31

Qy        940 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              ||||||||||||||||||||||||||||||
Db         30 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1


RESULT 19
AC143361
LOCUS       AC143361               161043 bp    DNA     linear   HTG 09-APR-2003
DEFINITION  Macaca mulatta clone CH250-271N12, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC143361
VERSION     AC143361.1  GI:29648548
KEYWORDS    HTG; HTGS_PHASE2; HTGS_PGI.
SOURCE      Macaca mulatta (rhesus monkey)
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 161043)
  AUTHORS   Csuros,M. and Milosavljevic,A.
  TITLE     Pooled genomic indexing (PGI): mathematical analysis and experiment
            design
  JOURNAL   (in) Guigo,R. and Gusfield,D. (Eds.);
            ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
            2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
            Springer (2002)
REFERENCE   2  (bases 1 to 161043)
  AUTHORS   Milosavljevic,A., Sodergren,E., Csuros,M., Li,B., Jackson,A.R.,
            Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L.,
            Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
```

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Egan,A., Earnhart,C., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 161043)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: LCGK
Center clone name: CH250-271N12
---------- Summary Statistics
Chemistry: Dye-primer Bodipy: inf% of reads
Chemistry: Dye-terminator Big Dye: inf% of reads
Consensus quality: 7124 bases at least Q40
Consensus quality: 8907 bases at least Q30
Consensus quality: 10743 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the PGI method using the Human genome (NCBI build 31)
* as the comparative genome.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 161043: contig of 161043 bp in length.

FEATURES Location/Qualifiers
source 1. .161043
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="CH250-271N12"
misc_feature 1. .161043
/note="assembly_name:CH250-271N12.1B
CONFIDENCE:_0.83"
BASE COUNT 2845 a 3627 c 3686 g 2810 t 148075 others
ORIGIN

Query Match 12.9%; Score 127.8; DB 2; Length 161043;
Best Local Similarity 76.3%; Pred. No. 4.2e-20;
Matches 135; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy        529 CTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCC 588
              | |                                 ||||||| || |||||| ||||| |||
Db          1 CGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGGTCCATCTCCAGCAGCGACCACGCC 60

Qy        589 ACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCC 648
              || |||||||||||||||||||||||||||||| ||||||||||||||||||  |||||
Db         61 ACGGGCAGCCTGCCAACTTCTCCTTCCTGCCGGGCCAGACATCGGACTGGTTCCAGTGCC 120

Qy        649 AGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTG 705
              |||||||||||| || |||||||||||||||||||||||| ||||||||||||||
Db        121 AGGCTGCAAACAGCGTCAATGTCCAGCACAGCGCCCTCACACTGGTGCCCCCAGGTG 177
```

RESULT 20
AC003666
LOCUS AC003666 151750 bp DNA linear PRI 07-AUG-1999
DEFINITION Homo sapiens Xp22 BAC GS-551O19 (Genome Systems Human BAC library) and cosmids U199A7 and U209F2 (Lawrence Livermore X chromosome cosmid library) containing part of human chloride channel 4 gene, complete sequence.
ACCESSION AC003666 AC002358 AC002360 AC003017
VERSION AC003666.1 GI:2992476
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151750)
AUTHORS Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,J., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L., Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J., Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151750)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 151750)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 151750)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 151750)

```
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 151750)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1. .151750
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp22"
                     /clone="GS-551O19, U199A7, U209F2"
                     /clone_lib="Genome Systems Human BAC library, Lawrence
                     Livermore X chromosome library"
     misc_feature    1. .374
                     /note="Overlaps bases 166883 to 167256 in AC008008"
                     /function="Overlap with clone AC008008"
     repeat_region   complement(2657. .2689)
                     /rpt_family="AT_rich"
     repeat_region   complement(2773. .2863)
                     /rpt_family="MER5B"
     repeat_region   3472. .4871
                     /rpt_family="L1PA2"
     repeat_region   5843. .6143
                     /rpt_family="L1MA8"
     repeat_region   complement(6989. .7493)
                     /rpt_family="L1MA3"
     repeat_region   complement(7513. .7811)
                     /rpt_family="AluSx"
     repeat_region   11128. .11201
                     /rpt_family="(GGAA)n"
     STS             11487. .11618
                     /standard_name="hz-1"
     repeat_region   11757. .11981
                     /rpt_family="MER20"
     repeat_region   12660. .12782
                     /rpt_family="L2"
     repeat_region   12879. .13195
                     /rpt_family="MLT1E"
     repeat_region   complement(15364. .15659)
                     /rpt_family="AluJo"
     repeat_region   complement(15836. .16099)
                     /rpt_family="MER20"
     repeat_region   complement(16717. .17085)
                     /rpt_family="L1MB7"
     repeat_region   complement(17159. .17374)
                     /rpt_family="L1MB7"
     repeat_region   complement(17421. .19544)
                     /rpt_family="L1M4"
     repeat_region   complement(19654. .19760)
                     /rpt_family="L1MB7"
     repeat_region   complement(20189. .20669)
                     /rpt_family="L1MB7"
     repeat_region   20670. .21171
                     /rpt_family="MLT1F"
     repeat_region   complement(21172. .21792)
                     /rpt_family="L1MB7"
     repeat_region   22005. .22430
                     /rpt_family="L1MC1"
     repeat_region   23332. .23445
                     /rpt_family="L1PA4"
     repeat_region   23469. .23588
                     /rpt_family="FLAM_A"
     repeat_region   complement(25042. .25067)
                     /rpt_family="POLY_A"
     repeat_region   complement(25163. .25265)
                     /rpt_family="L1PA5"
     repeat_region   25264. .25431
                     /rpt_family="L1PA5"
     repeat_region   complement(25569. .26063)
                     /rpt_family="MER31A"
     repeat_region   complement(27469. .27570)
                     /rpt_family="MIR"
     repeat_region   28608. .28643
                     /rpt_family="(TAAAA)n"
     repeat_region   complement(28713. .29010)
                     /rpt_family="AluJb"
     repeat_region   29209. .29310
                     /rpt_family="MIR"
     repeat_region   29312. .29613
                     /rpt_family="AluSx"
     repeat_region   30127. .30683
                     /rpt_family="MLT1D"
     repeat_region   31072. .31160
                     /rpt_family="(GAA)n"
     repeat_region   31161. .31279
                     /rpt_family="(GGA)n"
     repeat_region   complement(31772. .31891)
                     /rpt_family="(GA)n"
     repeat_region   32352. .32572
                     /rpt_family="MIR"
     repeat_region   complement(33455. .33479)
                     /rpt_family="AT_rich"
     repeat_region   34912. .35382
                     /rpt_family="L1MA1"
     repeat_region   35405. .35701
                     /rpt_family="AluSg"
     repeat_region   35707. .36176
                     /rpt_family="L1MA1"
     repeat_region   36178. .38428
                     /rpt_family="L1PA2"
     STS             36898. .37069
                     /standard_name="STS1-cSRL-31b6-uA/cSRL-31b6-uZ, Chr. -,
                     Homo sapiens"
                     /db_xref="dbSTS:7821"
     STS             37958. .38089
                     /standard_name="A002D07, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:32826"
     repeat_region   38436. .38566
                     /rpt_family="L1MA1"
     repeat_region   38597. .38795
                     /rpt_family="L1PB3"
     repeat_region   complement(38769. .40304)
                     /rpt_family="L1PA12"
     repeat_region   40309. .40609
                     /rpt_family="AluY"
     repeat_region   complement(40626. .40905)
                     /rpt_family="L1PA12"
     repeat_region   40904. .41028
                     /rpt_family="L1PB3"
     repeat_region   complement(41977. .42037)
                     /rpt_family="AT_rich"
     repeat_region   complement(42391. .42718)
                     /rpt_family="MER34"
```

```
     repeat_region   42979. .43067
                     /rpt_family="L2"
     repeat_region   complement(43085. .43220)
                     /rpt_family="MIR"
     repeat_region   43446. .43746
                     /rpt_family="AluSx"
     repeat_region   44344. .44380
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(45352. .45584)
                     /rpt_family="MIR"
     repeat_region   complement(46643. .46707)
                     /rpt_family="L2"
     repeat_region   complement(46905. .47293)
                     /rpt_family="MSTA"

  Query Match             7.0%;  Score 69.2;  DB 9;  Length 151750;
  Best Local Similarity   66.8%;  Pred. No. 5.7e-06;
  Matches  131;  Conservative    0;  Mismatches   58;  Indels    7;  Gaps    2;

Qy        796 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 855
              |||||||     ||    |||   || |   |||||||||||||  | ||| |||||  |
Db       5846 AGGAGTTAACCAGGGGTGGGAATAGGGAGATGGAGGTCAGAGGATACGAAGGAGCAGGTA 5905

Qy        856 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCA---GAGGACTGCAGGCCATCAGCGT 912
              |||||  ||||||    | |     | | |||  ||   |||||| || | ||| |   |
Db       5906 TGTAGGATGAACAAATCTGGAGATCTAACACACCACCTGAGGACTACATGGCATAAAATT 5965

Qy        913 GCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAA 972
                |||     ||||||  |||||   ||||||||   ||||||||||||||||||||
Db       5966 ATACTGA----ATTTGGGATTCATAAGAAATGAGTAGATTTTAGCTGCTCTTGCCACAAA 6021

Qy        973 AAAAAAAAAAAAAAAA 988
              || |||||  ||||||
Db       6022 AATAAAAACGAAAAAA 6037


RESULT 21
AC103588
LOCUS       AC103588               208670 bp    DNA     linear   PRI 21-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-640H5, complete sequence.
ACCESSION   AC103588
VERSION     AC103588.2  GI:19570161
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 208670)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 208670)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-NOV-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 208670)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 21, 2002 this sequence version replaced gi:17149452.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-640H5 (bc0507)
            ------------- Summary Statistics
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 208594 bases at least Q40
                Consensus quality: 208666 bases at least Q30
                Consensus quality: 208670 bases at least Q20
                Insert size: 207271; sum-of-contigs
                Quality coverage: 7.7x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5': Mapping in progress
                3': RP11-1C5 (UWGC:bc0101) AC093411, 107464-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                   EcoRI                    BglII                   HindIII
            ------------------  ------------------  ------------------
            SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
            ---------  --------  ---------  --------  ---------  --------

            ---------             ---------            ---------
                 8696      8812       2524      2527        561       <800
            ---------             ---------            ---------
                    6      <800       2067      2075       6382       6445
            ---------             ---------            ---------
                 1516      1544       9401      9533        512       <800
            ---------             ---------            ---------
                 5561      5606      17986     18248        449       <800
            ---------             ---------            ---------
                10971     10860       4692      4675       2484       2483
            ---------             ---------            ---------
                 2516      2558       2347      2350      11918      12069
            ---------             ---------            ---------
                 1834      1834       7017      7123        511       <800
            ---------             ---------            ---------
                 2909      2978       3961      3953       1038       1087
            ---------             ---------            ---------
                 1008      1008       1756      1735       4818       4777
            ---------             ---------            ---------
                 2704      2761       6058      5990       1677       1673
            ---------             ---------            ---------
                13906     13758       5740      5669       2657       2610
            ---------             ---------            ---------
                 1875      1834       4526      4503       1348       1312
            ---------             ---------            ---------
                 3723      3725       3178      3199        462       <800
```

2292	2308	773	<800	2481	2483
2981	2978	1639	1735	88	<800
2132	2153	1523	1485	1731	1673
8772	8812	1514	1485	4039	3956
594	<800	62	<800	2376	2365
4622	4505	1698	1735	96	<800
7761	7875	1776	1735	2802	2806
16778	17016	3527	3557	1989	1860
1849	1834	7630	7751	651	<800
4534	4505	919	860	718	<800
368	<800	922	934	9636	9551
2422	2446	197	<800	2594	2610
1552	1544	847	806	239	<800
2280	2308	4549	4503	462	<800
6530	6504	232	<800	3573	3584
2532	2558	1546	1560	3228	3296
8980	8812	1987	1938	3208	3185
11726	11389	3811	3782	1065	1087
178	<800	7829	7751	1512	1474
8243	8232	135	<800	1711	1673
7944	7875	422	<800	580	<800
1983	1983	38	<800	1599	1673
8839	8812	3158	3199	2377	2365
2269	2153	3675	3782	446	<800
9343	9320	2325	2350	2793	2806
4787	4779	7858	7751	1695	1673
11	<800	1958	1938	5164	5013
205	<800	11075	10794	9217	9166
3339	3348	7157	7123	312	<800
2822	2852	1909	1938	8581	8486
9454	9320	207	<800	12571	12653
3920	3934	1341	1344	1513	1474
1572	1544	10088	10090	2661	2610
6530	6504	1179	1213	5588	5507
		1257	1213	659	<800
		3777	3782	233	<800

10346	10090	1901	1860
10998	10794	275	<800
267	<800	2573	2610
715	<800	6470	6445
4942	4854	2385	2365
1034	1043	1119	1087
3805	3782	3192	3074
7558	7751	5056	5013
4678	4675	21	<800
1233	1213	4314	4314

Query Match 7.0%; Score 69.2; DB 9; Length 208670;
Best Local Similarity 63.3%; Pred. No. 5.7e-06;
Matches 124; Conservative 0; Mismatches 68; Indels 4; Gaps 1;

```
Qy        795 GAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCC 854
              ||  | |||||  ||    || |  ||| |   | | |||||||||  ||||| |||||
Db      89973 GATTACTTTGGATGGGGAAGAAATGGGGAGATGTAAGTCAGAGGATACAAAGTAGCAGAT 90032

Qy        855 ATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGC 914
              || ||| |||| |      |          | |  | |  | | ||||| ||   || |   ||
Db      90033 ATATAGGATGAGCACATCTAGAGATCTAATATGCAACATGACTGTAGTTAATAAAATTGT 90092

Qy        915 ACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
              |||    |||||||||  || |||| ||| || |   |||||||||| ||||||| ||||
Db      90093 ACT----GTATTTGGAATTTATGCTAAAAGAATAGATTTTAGCTGCCCTTGCCATAAAAG 90148

Qy        975 AAAAAAAAAAAAAAAA 990
              ||| ||| ||||| |
Db      90149 CAAACAAACAAAAATA 90164
```

RESULT 22
AL353752
LOCUS AL353752 177921 bp DNA linear PRI 25-MAY-2001
DEFINITION Human DNA sequence from clone RP11-497K15 on chromosome 9, complete sequence.
ACCESSION AL353752
VERSION AL353752.6 GI:13751335
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 177921)
AUTHORS Tracey,A.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Apr 21, 2001 this sequence version replaced gi:13445347.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-497K15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-497K15 The true left end of clone RP11-40C6 is at 128837 in this sequence. The true right end of clone RP11-58O2 is at 80971 in this sequence.

FEATURES Location/Qualifiers

source 1. .177921
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-497K15"
/clone_lib="RPCI-11.2"

misc_feature 7. .372
/note="match: GSS: Em:AQ521699"

misc_feature 7. .187
/note="match: GSS: Em:AQ400784"

repeat_region 507. .633
/note="L2 repeat: matches 2370. .2494 of consensus"

repeat_region 649. .751
/note="L1MB8 repeat: matches 6074. .6173 of consensus"

repeat_region 1133. .1428
/note="AluSx repeat: matches 1. .297 of consensus"

repeat_region 3216. .3644
/note="MSTB repeat: matches 1. .420 of consensus"

repeat_region 6009. .6477
/note="MLT1D repeat: matches 3. .501 of consensus"

repeat_region 6702. .6982
/note="AluJb repeat: matches 1. .308 of consensus"

repeat_region 7119. .7203
/note="MIR repeat: matches 81. .167 of consensus"

misc_feature 7337. .8076
/note="match: GSS: Em:AQ752801"

repeat_region 8330. .8613
/note="L2 repeat: matches 2465. .2750 of consensus"

repeat_region 8633. .8927
/note="L2 repeat: matches 2088. .2408 of consensus"

repeat_region 9107. .9266
/note="L2 repeat: matches 2593. .2750 of consensus"

repeat_region 9700. .10001
/note="AluSc repeat: matches 1. .303 of consensus"

repeat_region 10683. .10783
/note="MIR repeat: matches 31. .145 of consensus"

repeat_region 10890. .10969
/note="MIR repeat: matches 81. .159 of consensus"

misc_feature complement(11756. .12346)
/note="match: STS: Em:HSJ20A4"

repeat_region 11931. .11992
/note="L2 repeat: matches 2688. .2750 of consensus"

repeat_region 12476. .12860
/note="MLT1J repeat: matches 106. .514 of consensus"

repeat_region 13210. .13318
/note="FLAM_A repeat: matches 1. .109 of consensus"

misc_feature complement(13812. .14276)
/note="match: GSS: Em:B58243"

repeat_region 13898. .14064
/note="FRAM repeat: matches 0. .168 of consensus"

repeat_region 14484. .14783
/note="AluSx repeat: matches 1. .301 of consensus"

misc_feature 16037. .16512
/note="match: GSS: Em:AQ515148"

misc_feature 16045. .16482
/note="match: GSS: Em:AQ131836"

repeat_region 16114. .16288
/note="AluYa5 repeat: matches 133. .307 of consensus"

repeat_region 16646. .16717
/note="MLT1J repeat: matches 104. .187 of consensus"

repeat_region 16802. .17084
/note="L1MB8 repeat: matches 5878. .6173 of consensus"

repeat_region 17504. .17660
/note="L1M4 repeat: matches 5296. .5456 of consensus"

repeat_region 17657. .18355
/note="L1ME3 repeat: matches 5349. .6046 of consensus"

repeat_region 18357. .18656
/note="AluSp repeat: matches 1. .302 of consensus"

repeat_region 19964. .20259
/note="AluSq repeat: matches 5. .299 of consensus"

repeat_region 20604. .20912
/note="L2 repeat: matches 302. .628 of consensus"

repeat_region 20983. .21292
/note="AluY repeat: matches 1. .309 of consensus"

repeat_region 21321. .21428
/note="FLAM_A repeat: matches 21. .128 of consensus"

repeat_region 22740. .22830
/note="MIR repeat: matches 30. .125 of consensus"

repeat_region 22890. .22940
/note="L2 repeat: matches 2700. .2749 of consensus"

misc_feature complement(23361. .23469)
/note="match: STS: Em:U67307"

repeat_region 24070. .24362
/note="AluSq repeat: matches 6. .296 of consensus"

repeat_region 24403. .25216
/note="L1MA6 repeat: matches 5487. .6295 of consensus"

repeat_region 25217. .25355
/note="L1PB1 repeat: matches 6011. .6151 of consensus"

repeat_region 25356. .25420
/note="L1MA6 repeat: matches 5425. .5487 of consensus"

repeat_region 25415. .25744
/note="L1MA6 repeat: matches 5091. .5403 of consensus"

repeat_region 26967. .27168
/note="MIR repeat: matches 49. .262 of consensus"

repeat_region 27190. .27327
/note="Charlie1 repeat: matches 2620. .2761 of consensus"

repeat_region 27323. .27475
/note="Charlie1 repeat: matches 64. .226 of consensus"

repeat_region 27607. .27917
/note="AluSx repeat: matches 1. .306 of consensus"

repeat_region 28407. .29843
/note="L1MB7 repeat: matches 4591. .6173 of consensus"

repeat_region 32140. .32448
/note="AluSx repeat: matches 1. .304 of consensus"

repeat_region 33717. .34022
/note="AluSc repeat: matches 1. .306 of consensus"

repeat_region 34170. .34419
/note="MLT1G repeat: matches 31. .260 of consensus"

repeat_region 34527. .34626
/note="MLT1F repeat: matches 440. .541 of consensus"

repeat_region 35085. .35128
/note="BUR1 repeat: matches 4967. .5010 of consensus"

repeat_region 35525. .35836
/note="AluSx repeat: matches 1. .312 of consensus"

repeat_region 35844. .35911
/note="MLT1I repeat: matches 349. .409 of consensus"

repeat_region 36507. .36760
/note="L1PB2 repeat: matches 4128. .4385 of consensus"

repeat_region 36761. .37068
/note="AluY repeat: matches 1. .308 of consensus"

repeat_region 37069. .38794
/note="L1PB2 repeat: matches 4385. .6105 of consensus"

repeat_region 38795. .39098
/note="AluJb repeat: matches 2. .302 of consensus"

repeat_region 39099. .39146
/note="L1PB2 repeat: matches 6105. .6152 of consensus"

repeat_region 39612. .39895
/note="AluSx repeat: matches 1. .284 of consensus"

```
     repeat_region   40251. .40597
                     /note="L2 repeat: matches 2390. .2747 of consensus"
     repeat_region   40814. .41296
                     /note="MLT1D repeat: matches 1. .505 of consensus"
     repeat_region   43064. .43491
                     /note="MLT2FB repeat: matches 1. .412 of consensus"
     repeat_region   43800. .43934
                     /note="MLT1H repeat: matches 412. .543 of consensus"
     repeat_region   44369. .44772
                     /note="MSTA repeat: matches 3. .426 of consensus"
     repeat_region   45202. .45287
                     /note="MIR repeat: matches 115. .208 of consensus"
     repeat_region   45352. .45855
                     /note="MER1A repeat: matches 1. .523 of consensus"
     repeat_region   46156. .46517
                     /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region   47513. .47824
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   48695. .48996
                     /note="AluSc repeat: matches 1. .304 of consensus"
     repeat_region   49341. .49553
                     /note="MIR repeat: matches 6. .252 of consensus"
     misc_feature    complement(49807. .50346)
                     /note="match: GSS: Em:AQ174373"

  Query Match             6.3%;  Score 62.8;  DB 9;  Length 177921;
  Best Local Similarity   61.6%;  Pred. No. 0.0002;
  Matches  117;  Conservative    0;  Mismatches   72;  Indels    1;  Gaps    1;

Qy        802 TTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGA 861
              ||| |||         |  ||| |   | ||||||||||||| ||||| ||||| | |||||
Db      76569 TTGAGGGAAGAGCAAATGGGGAGATGTAGGTCAGAGGACACAAAGTAGCAGACGTGTAGG 76628

Qy        862 ATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTC 921
              ||||||    |    || | |       | | || | |  |   | | |      | |
Db      76629 ATGAACAAAACTAGAAGACTAATGTACAACATGATTTCTAGAGGTAATCAAATTGTACT- 76687

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
              |||||||  ||| ||| |||| |||   || ||||||||||||||| ||||  ||||
Db      76688 GTATTTGGGATTCCTGCCAAATTAGTAGATTCTAGCTGCTCTTGCCACAGAAAAGGAAAA 76747

Qy        982 AAAAAAAAAA 991
              |||   | ||
Db      76748 AAATGGATAA 76757


RESULT 23
AC018788/c
LOCUS       AC018788              146883 bp    DNA     linear   PRI 22-NOV-2001
DEFINITION  Homo sapiens chromosome , clone RP11-24E9, complete sequence.
ACCESSION   AC018788
VERSION     AC018788.10  GI:16117587
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-24E9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H.,
            O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A.,
            Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A.
            and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 146883)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 13, 2001 this sequence version replaced gi:16041478.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4443
                Center clone name: 24_E_9
            --------------.
FEATURES             Location/Qualifiers
     source          1. .146883
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-24E9"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(455. .723)
                     /rpt_family="MIR"
     repeat_region   2337. .2409
                     /rpt_family="L2"
     repeat_region   complement(2418. .2815)
                     /rpt_family="L3"
     repeat_region   complement(3143. .3415)
                     /rpt_family="L2"
     repeat_region   complement(3428. .3588)
                     /rpt_family="L3"
     repeat_region   4076. .4116
                     /rpt_family="MIR"
     repeat_region   complement(4701. .4992)
                     /rpt_family="AluSg"
     repeat_region   5007. .5205
                     /rpt_family="MIR"
     repeat_region   5300. .5371
                     /rpt_family="L2"
```

```
repeat_region   complement(7086. .7283)
                /rpt_family="MER5A"
repeat_region   7418. .7464
                /rpt_family="L2"
repeat_region   complement(8426. .8712)
                /rpt_family="L1PA10"
repeat_region   complement(8753. .8917)
                /rpt_family="MER54A"
repeat_region   complement(8984. .9076)
                /rpt_family="MIR"
repeat_region   9639. .9748
                /rpt_family="MIR"
repeat_region   complement(9766. .10614)
                /rpt_family="MER54A"
repeat_region   10700. .10834
                /rpt_family="L2"
repeat_region   10859. .10967
                /rpt_family="GA-rich"
repeat_region   complement(11035. .11307)
                /rpt_family="L3"
repeat_region   11316. .11349
                /rpt_family="(TC)n"
repeat_region   11349. .11391
                /rpt_family="(CA)n"
repeat_region   complement(11869. .11975)
                /rpt_family="MLT1G1"
repeat_region   complement(12252. .14205)
                /rpt_family="L1PA5"
repeat_region   complement(14965. .15444)
                /rpt_family="MER114"
repeat_region   15495. .15722
                /rpt_family="L1ME1"
repeat_region   15819. .16227
                /rpt_family="L1P3"
repeat_region   16228. .16256
                /rpt_family="(TTTTG)n"
repeat_region   16257. .16678
                /rpt_family="L1P3"
repeat_region   16679. .17746
                /rpt_family="MER11C"
repeat_region   17747. .19574
                /rpt_family="L1P3"
repeat_region   19575. .21107
                /rpt_family="L1PA7"
repeat_region   complement(24551. .24629)
                /rpt_family="MIR"
repeat_region   complement(24706. .25065)
                /rpt_family="L1PB1"
repeat_region   25066. .25181
                /rpt_family="(TA)n"
repeat_region   25183. .25866
                /rpt_family="L1"
repeat_region   25871. .25976
                /rpt_family="(TA)n"
repeat_region   complement(25991. .26108)
                /rpt_family="L1P5"
repeat_region   complement(26175. .26288)
                /rpt_family="L1PA15-16"
repeat_region   complement(26305. .26969)
                /rpt_family="L1PA15-16"
repeat_region   26964. .27157
                /rpt_family="L1"
repeat_region   complement(27164. .27230)
                /rpt_family="MER61D"
repeat_region   27296. .27400
                /rpt_family="L1PA16"
repeat_region   complement(27401. .27697)
                /rpt_family="AluSp"
repeat_region   27698. .29368
                /rpt_family="L1PA16"
unsure          29271. .29293
                /note="<30  qual SNGL region"
repeat_region   complement(29369. .29564)
                /rpt_family="MER61D"
repeat_region   29981. .30691
                /rpt_family="L1PA16"
repeat_region   31815. .31838
                /rpt_family="AT_rich"
repeat_region   32875. .32910
                /rpt_family="AT_rich"
repeat_region   32979. .33163
                /rpt_family="MLT1A1"
repeat_region   33164. .33465
                /rpt_family="AluJb"
repeat_region   33466. .33616
                /rpt_family="MLT1A1"
repeat_region   33704. .34140
                /rpt_family="MLT1K"
repeat_region   34745. .34829
                /rpt_family="MLT1K"
repeat_region   35312. .35700
                /rpt_family="L2"
repeat_region   36237. .37252
                /rpt_family="L1MC1"
repeat_region   complement(37557. .38067)
                /rpt_family="LTR40a"
repeat_region   39342. .39757
                /rpt_family="MLT2A2"
repeat_region   39760. .39818
                /rpt_family="(CATATA)n"
repeat_region   39819. .39874
                /rpt_family="MLT2A2"
```

Query Match 6.2%; Score 61.8; DB 9; Length 146883;
Best Local Similarity 57.5%; Pred. No. 0.00035;
Matches 111; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

```
Qy        796 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 855
              ||||| | | | ||   || |  ||| |   | ||||||||||  | |||  ||   |
Db     100174 AGGAGGTGGAGAGGGGGAGAAATGGGGAGATGAAGGTCAGAGGATACGAAGTTACAAATA 100115

Qy        856 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCA 915
              |||| ||| |||    | |   ||     ||     |    | |   |
Db     100114 TGTAAAATAAACAACTCTGGAGACCCCTATGTACACCATGAGGACTATGGTTAATAAGAT 100055

Qy        916 CTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAA 975
                ||||||||  ||  ||| |||||||| | |||||||| |||||||||| ||||
Db     100054 TGTACTGTATTTGGGATTTCTGCTAAATGAGTATATTTTAGCTTCTCTTGCCACCAAAAC 99995

Qy        976 AAAAAAAAAAAAA 988
              |||||  ||||||
Db      99994 AAAAATTAAAAAA 99982
```

```
RESULT 24
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="synonym: JP3"
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

  Query Match             6.1%;  Score 61;  DB 9;  Length 125020;
  Best Local Similarity   13.2%;  Pred. No. 0.00054;
  Matches   91;  Conservative  299;  Mismatches  287;  Indels   12;  Gaps    3;

Qy         11 AGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAGGCATGGGGCT 70
              |:  ::::::::: ::::    :|:   |:|::||:|  : : :|:|    :|: :: |:
Db      17095 AKWYMYYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCYAKCKSMCY 17154

Qy         71 CCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGA 130
              :|::| :::|::::  :: |::::: | :|:| :      ||| | |     |  || | |
Db      17155 SCYYGMSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNANTGGGGAA 17214

Qy        131 AGAAATTACCCCTGTGGTCTCCATTGCCTACA-------AAGTCCTGGAAGTTTTCCCCA 183
                    |      ::: ::   ||  ::   :        :|| :||:   : |  : ::
Db      17215 GCTTTTNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSMTMSMASYC 17274

Qy        184 AAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATT 243
              :  ::: |  ::  :  : : :: :: : : :::|:::||:|: :| |::: :::  : :
Db      17275 WCMSMYYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRSMRSRGMSY 17334

Qy        244 CCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGG 303
              ::: ::: ::: |:  :     :  : ::::::: : :: :| :: :   ::  |  : ::
Db      17335 YMYASWKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSMWGSASRSS 17394

Qy        304 CCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCC 363
              |::|:: :: :: :  |  ||  | :: | :::   |:::          ::::  :::
Db      17395 CYKCYKSMRCSMMSSKCYRCA-GCMMMKGGYMRYMRCWSMKRRWWGKSAMYRMRRMWKGR 17453

Qy        364 GGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGC 423
              |  :  :  ::::: |:   :::||   :  |:::  ::   :::|:: : | |:  :  :
Db      17454 GAMMMCMKCYSRMSRCMMMMKSYCASCRSCAMMSGTYMKCASYMCMSSYWCTCWCTSYMS 17513

Qy        424 TGTGGTCCAAGCCAGTGTCTGAGCTGCGGG----CCAACTTCACTCTGCAGGACAGAGGG 479
              :      |  ::  :::::: :: ::::|    ::: ||     :   :   ::||::
Db      17514 YYRCTCWCKGWSSYTKSKSWSSSMSSYKKGRRKSYSMCCTSRGAMSCWRRCCYMRGASSM 17573

Qy        480 GCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAAC 539
              :  : :   |:|:  | : : :   |: :::: ::|| :: :: ::| |  : : :: ::

Db      17574 RAGSMSRRRAKGRSWGGRSKWWMTGGMWRSKYYYYCTGRRMMMTYMCCCWRRRSYMYYRS 17633

Qy        540 AGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCT 599
              :: :  :::: : :: : : :::: ::   :  | : ::  : : ::: :  |:  ::
Db      17634 MSAMGMRKSSWSGMRMMGSASSRRCKSASRSSWCSRRMKGMRGSCWSSKMWGGSRSRSAS 17693

Qy        600 GCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAAC 659
              : |  :  :    : : :: ::|   ::::| :: ::|: :|::: |:: ::|: : | :
Db      17694 SKCKGSRGMRRRSKRSSKYRKRGRGKKRSMTKSKGSKGKSKCWKRSGSMTSSCYYYSASS 17753

Qy        660 AACGCCAATGTCCAGCACAGCGCCCTCAC 688
              :  ::     ||  :: ::| ||: :::
Db      17754 CMWMMSSKSCMCCCMMMKRRCACCYSSMS 17782


RESULT 25
AC016177
LOCUS       AC016177               164762 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 9 clone RP11-276N1 map 9, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC016177
VERSION     AC016177.4  GI:7533977
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164762)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 9, clone RP11-276N1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164762)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 164762)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
```

```
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 11, 2000 this sequence version replaced gi:7280308.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4198
                Center clone name: 276_N_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 148809 bases at least Q40
                Consensus quality: 157187 bases at least Q30
                Consensus quality: 160494 bases at least Q20
                Insert size: 168000; agarose-fp
                Insert size: 162462; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1218: contig of 1218 bp in length
            *     1219     1318: gap of 100 bp
            *     1319     2383: contig of 1065 bp in length
            *     2384     2483: gap of 100 bp
            *     2484     3871: contig of 1388 bp in length
            *     3872     3971: gap of 100 bp
            *     3972     4836: contig of 865 bp in length
            *     4837     4936: gap of 100 bp
            *     4937     6261: contig of 1325 bp in length
            *     6262     6361: gap of 100 bp
            *     6362     8332: contig of 1971 bp in length
            *     8333     8432: gap of 100 bp
            *     8433    11370: contig of 2938 bp in length
            *    11371    11470: gap of 100 bp
            *    11471    16136: contig of 4666 bp in length
            *    16137    16236: gap of 100 bp
            *    16237    20960: contig of 4724 bp in length
            *    20961    21060: gap of 100 bp
            *    21061    25287: contig of 4227 bp in length
            *    25288    25387: gap of 100 bp
            *    25388    32039: contig of 6652 bp in length
            *    32040    32139: gap of 100 bp
            *    32140    38768: contig of 6629 bp in length
            *    38769    38868: gap of 100 bp
            *    38869    43784: contig of 4916 bp in length
            *    43785    43884: gap of 100 bp
            *    43885    50401: contig of 6517 bp in length
            *    50402    50501: gap of 100 bp
            *    50502    57082: contig of 6581 bp in length
            *    57083    57182: gap of 100 bp
            *    57183    63917: contig of 6735 bp in length
            *    63918    64017: gap of 100 bp
            *    64018    71454: contig of 7437 bp in length
            *    71455    71554: gap of 100 bp
            *    71555    79670: contig of 8116 bp in length
            *    79671    79770: gap of 100 bp
            *    79771    87413: contig of 7643 bp in length
            *    87414    87513: gap of 100 bp
            *    87514    93773: contig of 6260 bp in length
            *    93774    93873: gap of 100 bp
            *    93874   103590: contig of 9717 bp in length
            *   103591   103690: gap of 100 bp
            *   103691   118309: contig of 14619 bp in length
            *   118310   118409: gap of 100 bp
            *   118410   136058: contig of 17649 bp in length
            *   136059   136158: gap of 100 bp
            *   136159   164762: contig of 28604 bp in length.
FEATURES                Location/Qualifiers
     source             1. .164762
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="9"
                        /map="9"
                        /clone="RP11-276N1"
                        /clone_lib="RPCI-11 Human Male BAC"
     misc_feature       1. .1218
                        /note="assembly_fragment"
     misc_feature       1319. .2383
                        /note="assembly_fragment"
     misc_feature       2484. .3871
                        /note="assembly_fragment"
     misc_feature       3972. .4836
                        /note="assembly_fragment
                        clone_end:SP6
                        vector_side:right"
     misc_feature       4937. .6261
                        /note="assembly_fragment"
     misc_feature       6362. .8332
                        /note="assembly_fragment"
     misc_feature       8433. .11370
                        /note="assembly_fragment"
     misc_feature       11471. .16136
                        /note="assembly_fragment"
     misc_feature       16237. .20960
                        /note="assembly_fragment"
     misc_feature       21061. .25287
                        /note="assembly_fragment"
     misc_feature       25388. .32039
                        /note="assembly_fragment"
     misc_feature       32140. .38768
                        /note="assembly_fragment"
     misc_feature       38869. .43784
                        /note="assembly_fragment"
     misc_feature       43885. .50401
                        /note="assembly_fragment"
     misc_feature       50502. .57082
                        /note="assembly_fragment"
     misc_feature       57183. .63917
                        /note="assembly_fragment"
     misc_feature       64018. .71454
                        /note="assembly_fragment"
     misc_feature       71555. .79670
                        /note="assembly_fragment"
     misc_feature       79771. .87413
                        /note="assembly_fragment"
     misc_feature       87514. .93773
                        /note="assembly_fragment
                        clone_end:T7
                        vector_side:left"
     misc_feature       93874. .103590
                        /note="assembly_fragment"
     misc_feature       103691. .118309
                        /note="assembly_fragment"
     misc_feature       118410. .136058
                        /note="assembly_fragment"
     misc_feature       136159. .164762
                        /note="assembly_fragment"
BASE COUNT    47298 a  33699 c  32793 g  48671 t   2301 others
ORIGIN

  Query Match             6.1%;  Score 60.8;  DB 2;  Length 164762;
  Best Local Similarity   64.1%;  Pred. No. 0.0006;
```

```
  Matches  109;  Conservative    0;  Mismatches   57;  Indels    4;  Gaps    1;

Qy        814 GGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCA 873
              | |  ||| |   | ||||||||||||| ||||| | ||| ||||||| ||||||    |
Db      64641 GAAATGGGGAGATGTAGGTCAGAGGACACAAAGTAACAGACATGTAGGATGAACAAGTCT 64700

Qy        874 GAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTT 933
                 |||         || || | |||   || |   || |||    |||||||   ||
Db      64701 AGAGATCAAATGTACGTGAAGATTACAGATAATAAAATTGTACT----GTATTTGTGATT 64756

Qy        934 CATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAA 983
              | ||| |||||||    |||||||  |||||||||||||  || ||||
Db      64757 CCTGCTAAATGAGTAGACTTTAGCTATTCTTGCCACAAAAACTAACAAAA 64806


RESULT 26
AL353764/c
LOCUS       AL353764              170898 bp    DNA     linear   PRI 01-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-440G5 on chromosome 9, complete
            sequence.
ACCESSION   AL353764
VERSION     AL353764.9  GI:14272263
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170898)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 31, 2001 this sequence version replaced gi:13990016.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-440G5 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-440G5 The true
            left end of clone RP11-8I8 is at 105684 in this sequence. The true
            right end of clone RP11-446A5 is at 70776 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .170898
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-440G5"
                     /clone_lib="RPCI-11.2"
BASE COUNT    48782 a  34836 c  35124 g  52156 t
ORIGIN
```

```
  Query Match             6.1%;  Score 60.8;  DB 9;  Length 170898;
  Best Local Similarity   64.1%;  Pred. No. 0.0006;
  Matches  109;  Conservative    0;  Mismatches   57;  Indels    4;  Gaps    1;

Qy        814 GGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCA 873
              | |  ||| |   | ||||||||||||| ||||| | ||| ||||||| ||||||    |
Db      69053 GAAATGGGGAGATGTAGGTCAGAGGACACAAAGTAACAGACATGTAGGATGAACAAGTCT 68994

Qy        874 GAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTT 933
                 |||         || || | |||   || |   || |||    |||||||   ||
Db      68993 AGAGATCAAATGTACGTGAAGATTACAGATAATAAAATTGTACT----GTATTTGTGATT 68938

Qy        934 CATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAA 983
              | ||| |||||||    |||||||  |||||||||||||  || ||||
Db      68937 CCTGCTAAATGAGTAGACTTTAGCTATTCTTGCCACAAAAACTAACAAAA 68888


RESULT 27
AC104684
LOCUS       AC104684              129770 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-1N7 from 2, complete sequence.
ACCESSION   AC104684 AC021641
VERSION     AC104684.3  GI:18464246
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129770)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 129770)
  AUTHORS   Belter,E., Meyer,R. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-1N7
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 129770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 129770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 129770)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18250104.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0001N07
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is XXcos-2052C30; the clone sequenced to the right is RP11-356M6, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-1N7.

Sequence derived from one plasmid subclone, base position 94919 to 94957.

Data from AC067919 and AC097644 was used to finish this clone, AC104684.

Polymorphisms have been identified between AC067919, AC079779 and AC104684.

The sequence of AC021641 has been incorporated into AC104684.

FEATURES Location/Qualifiers
source 1. .129770
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-1N7"
/clone_lib="RPCI-11"
repeat_region 1. .77
/rpt_family="L1"
repeat_region 78. .340
/rpt_family="Alu"
repeat_region 341. .668
/rpt_family="L1"
repeat_region 661. .932
/rpt_family="L1"
repeat_region 944. .1216
/rpt_family="Alu"
repeat_region 1231. .1277
/rpt_family="(TAAA)n"
repeat_region 1260. .1341
/rpt_family="L1"
repeat_region 1342. .1566
/rpt_family="Alu"
repeat_region 1567. .1624
/rpt_family="L1"
repeat_region 1615. .1636
/rpt_family="AT_rich"
repeat_region 1651. .1778
/rpt_family="Alu"
repeat_region 3395. .3458
/rpt_family="AcHobo"
repeat_region 3459. .3749
/rpt_family="Alu"
repeat_region 3750. .3801
/rpt_family="AcHobo"
repeat_region 3808. .3939
/rpt_family="MIR"
repeat_region 8004. .8034
/rpt_family="AT_rich"
repeat_region 8128. .8236
/rpt_family="CR1"
repeat_region 10147. .10658
/rpt_family="MaLR"
repeat_region 11021. .11493
/rpt_family="MaLR"
repeat_region 11851. .12002
/rpt_family="ERV1"
repeat_region 12684. .12975
/rpt_family="Alu"
repeat_region 13127. .13247
/rpt_family="MaLR"
repeat_region 13583. .13678
/rpt_family="ERV1"
repeat_region 13679. .14042
/rpt_family="ERV1"
repeat_region 15284. .15648
/rpt_family="MaLR"
repeat_region 17182. .17546
/rpt_family="MaLR"
repeat_region 18734. .18848
/rpt_family="MIR"
repeat_region 18851. .19169
/rpt_family="ERV1"
repeat_region 19405. .19541
/rpt_family="MIR"
repeat_region 19818. .20175
/rpt_family="MaLR"
repeat_region 21436. .21561
/rpt_family="L1"
repeat_region 21587. .22109
/rpt_family="ERV1"
repeat_region 23724. .23845
/rpt_family="MIR"
repeat_region 23925. .23987
/rpt_family="MIR"
repeat_region 24193. .24232
/rpt_family="L2"
repeat_region 24255. .24371
/rpt_family="MIR"
repeat_region 25991. .26298
/rpt_family="Alu"
repeat_region 26273. .26299
/rpt_family="AT_rich"
repeat_region 26380. .26415
/rpt_family="AT_rich"
repeat_region 26417. .26752
/rpt_family="L1"
repeat_region 26754. .27066
/rpt_family="L1"
repeat_region 27067. .27349
/rpt_family="Alu"
repeat_region 27350. .27667
/rpt_family="L1"
repeat_region 27457. .27494
/rpt_family="(CA)n"
repeat_region 28171. .28277
/rpt_family="(TA)n"
repeat_region 28336. .28384
/rpt_family="AT_rich"
repeat_region 28379. .28554
/rpt_family="L1"
repeat_region 28916. .29170
/rpt_family="L1"
repeat_region 29189. .29500
/rpt_family="L1"
repeat_region 29511. .29749
/rpt_family="L1"

```
     repeat_region   29751. .30738

  Query Match             6.1%;  Score 60.4;  DB 9;  Length 129770;
  Best Local Similarity   56.6%;  Pred. No. 0.00076;
  Matches  112;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        794 AGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGC 853
              || ||    | | | || |  ||   |||||     |||||||| |   ||||| |||||
Db      77435 AGGGGGACTGGAGAGGGTGGGGGATGGGAAGATATAGGTCAGAAGGTACAAAGTAGCAGG 77494

Qy        854 CATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTG 913
              ||| | || |||    |        |          | | ||   |  |   | |
Db      77495 TGTGTGGGATAAACAAGTCCAGAGATCTCATGTACAACATGAGGACTAGTGTTAAAAAAA 77554

Qy        914 CACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAA 973
                  |||||||| ||| ||  ||||||||   |||||| |||| || ||||||||
Db      77555 ATTGTGCTGTATTTGGAATTCCTGATAAATGAGTTGATTTTAGTTGCTATTACCACAAAA 77614

Qy        974 AAAAAAAAAAAAAAAAAA 991
                |||||  |||||  |||||
Db      77615 CAAAAACAAAAACAAAAA 77632


RESULT 28
AC067919/c
LOCUS       AC067919               171655 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-292I6, WORKING DRAFT SEQUENCE, 20 unordered
            pieces.
ACCESSION   AC067919
VERSION     AC067919.2  GI:8076830
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-292I6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7652358.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6333
                Center clone name: 292_I_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161528 bases at least Q40
                Consensus quality: 167303 bases at least Q30
                Consensus quality: 168939 bases at least Q20
                Insert size: 173000; agarose-fp
                Insert size: 169755; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      220: contig of 220 bp in length
            *      221      320: gap of 100 bp
            *      321     4001: contig of 3681 bp in length
            *     4002     4101: gap of 100 bp
            *     4102     7945: contig of 3844 bp in length
            *     7946     8045: gap of 100 bp
            *     8046    11450: contig of 3405 bp in length
            *    11451    11550: gap of 100 bp
            *    11551    16117: contig of 4567 bp in length
            *    16118    16217: gap of 100 bp
            *    16218    20480: contig of 4263 bp in length
            *    20481    20580: gap of 100 bp
            *    20581    25760: contig of 5180 bp in length
            *    25761    25860: gap of 100 bp
            *    25861    31039: contig of 5179 bp in length
            *    31040    31139: gap of 100 bp
            *    31140    35820: contig of 4681 bp in length
            *    35821    35920: gap of 100 bp
            *    35921    42490: contig of 6570 bp in length
            *    42491    42590: gap of 100 bp
            *    42591    49780: contig of 7190 bp in length
            *    49781    49880: gap of 100 bp
            *    49881    57796: contig of 7916 bp in length
            *    57797    57896: gap of 100 bp
            *    57897    67541: contig of 9645 bp in length
            *    67542    67641: gap of 100 bp
            *    67642    76045: contig of 8404 bp in length
            *    76046    76145: gap of 100 bp
            *    76146    85834: contig of 9689 bp in length
            *    85835    85934: gap of 100 bp
            *    85935    95701: contig of 9767 bp in length
            *    95702    95801: gap of 100 bp
            *    95802   107607: contig of 11806 bp in length
            *   107608   107707: gap of 100 bp
            *   107708   123352: contig of 15645 bp in length
            *   123353   123452: gap of 100 bp
            *   123453   141544: contig of 18092 bp in length
            *   141545   141644: gap of 100 bp
            *   141645   171655: contig of 30011 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171655
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-292I6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .220
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
```

```
     misc_feature    321. .4001
                     /note="assembly_fragment"
     misc_feature    4102. .7945
                     /note="assembly_fragment"
     misc_feature    8046. .11450
                     /note="assembly_fragment"
     misc_feature    11551. .16117
                     /note="assembly_fragment"
     misc_feature    16218. .20480
                     /note="assembly_fragment"
     misc_feature    20581. .25760
                     /note="assembly_fragment"
     misc_feature    25861. .31039
                     /note="assembly_fragment"
     misc_feature    31140. .35820
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    35921. .42490
                     /note="assembly_fragment"
     misc_feature    42591. .49780
                     /note="assembly_fragment"
     misc_feature    49881. .57796
                     /note="assembly_fragment"
     misc_feature    57897. .67541
                     /note="assembly_fragment"
     misc_feature    67642. .76045
                     /note="assembly_fragment"
     misc_feature    76146. .85834
                     /note="assembly_fragment"
     misc_feature    85935. .95701
                     /note="assembly_fragment"
     misc_feature    95802. .107607
                     /note="assembly_fragment"
     misc_feature    107708. .123352
                     /note="assembly_fragment"
     misc_feature    123453. .141544
                     /note="assembly_fragment"
     misc_feature    141645. .171655
                     /note="assembly_fragment"
BASE COUNT    50815 a  35321 c  34768 g  48847 t   1904 others
ORIGIN

  Query Match             6.1%;  Score 60.4;  DB 2;  Length 171655;
  Best Local Similarity   56.6%;  Pred. No. 0.00075;
  Matches  112;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        794 AGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGC 853
              || ||   | | | || |  ||   |||||     |||||||| |   ||||| |||||
Db     150859 AGGGGGACTGGAGAGGGTGGGGGATGGGAAGATATAGGTCAGAAGGTACAAAGTAGCAGG 150800

Qy        854 CATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTG 913
              ||| | || |||    |        |         | | ||   | |   | |
Db     150799 TGTGTGGGATAAACAAGTCCAGAGATCTCATGTACAACATGAGGACTAGTGTTAAAAAAA 150740

Qy        914 CACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAA 973
              |||||||| ||| ||  ||||||||   |||||| |||| || ||||||||
Db     150739 ATTGTGCTGTATTTGGAATTCCTGATAAATGAGTTGATTTTAGTTGCTATTACCACAAAA 150680

Qy        974 AAAAAAAAAAAAAAAAAA 991
              |||||  |||||  |||||
Db     150679 CAAAAACAAAAACAAAAA 150662


RESULT 29
AC010072
LOCUS       AC010072               124347 bp    DNA     linear   PRI 19-NOV-1999
DEFINITION  Homo sapiens chromosome 14q31 clone CTD-2173I4 containing TSHR
            gene, partial cds; and unknown gene, complete sequence.
ACCESSION   AC010072
VERSION     AC010072.5  GI:6453843
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124347)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Sequencing of human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124347)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
            Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 124347)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
            Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-NOV-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Nov 19, 1999 this sequence version replaced gi:6114900.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Big Dye terminators and primers
                Assembly program: Phrap; version 0.990399
            ------------.
FEATURES             Location/Qualifiers
     source          1. .124347
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q31"
                     /clone="CTD-2173I4"
                     /clone_lib="CalTech human BAC library D"
                     /note="This clone overlaps RP11-114N19, Accession
                     AC007262"
     CDS             complement(join(42575. .42673,43714. .43786,47934. .48025,
                     52809. .52927,53951. .54077,62318. .62371))
                     /note="Intron-exon boundaries defined in relation to EST
                     AI554064.  the closest BLASTX similarity is to a
                     drosophila potassium channel protein"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAF09033.1"
                     /db_xref="GI:6453845"
                     /translation="MLGRYREYSNGQAGAIEHLKESLEQSIDQLRSQRLLRNSGGRSI
                     SVTSLSASDLDGGTGSELHHFPPTSPLKDYGDPQGIKRMRSRTGVRFVQETDDMTQLH
                     GFHQSLRDLSSEQIRLGDDFNRELSRRSRSDAETKRALEELTEKLNEAQKQEVFFVKS
                     LGKVKIQCLSSAEVDNVFPINVWQCGI"
     unsure          95810. .95860
                     /note="low quality data"
     mRNA            103634. .>103846
                     /product="thyroid stimulating hormone receptor"
                     /note="This is the 5' end of TSHR, found in M73745"
     CDS             103677. .>103846
                     /note="thyroid stimulating hormone receptor"
                     /codon_start=1
                     /product="TSHR"
                     /protein_id="AAF09032.1"
                     /db_xref="GI:6453844"
                     /translation="MRPADLLQLVLLLDLPRDLGGMGCSSPPCECHQEEDFRVTCKDI
                     QRIPSLPPSTQTL"
```

```
     misc_feature    109619. .124347
                     /note="Overlap with RP11-114N19, Accession AC007262"
     variation       110216
                     /note="2173I4: c; 114N19: t"
                     /replace="t"
     variation       110755
                     /note="2173I4: g; 114N19: a"
                     /replace="a"
     variation       111115
                     /note="2173I4: c; 114N19: g"
                     /replace="g"
     variation       114198
                     /note="2173I4: g; 114N19: a"
                     /replace="a"
     variation       114690
                     /note="2173I4: c; 114N19: t"
                     /replace="t"
     variation       115183
                     /note="2173I4: t; 114N19: tatat"
                     /replace="tatat"
     unsure          115210. .115280
                     /note="low quality data"
     variation       116847
                     /note="2173I4: t; 114N19: c"
                     /replace="c"
     variation       117131
                     /note="2173I4: g; 114N19: a"
                     /replace="a"
     variation       117408
                     /note="2173I4: a; 114N19: c"
                     /replace="c"
     variation       117633
                     /note="2173I4: g; 114N19: t"
                     /replace="t"
     variation       118901
                     /note="2173I4: a; 114N19: g"
                     /replace="g"
     variation       118988
                     /note="2173I4: c; 114N19: t"
                     /replace="t"
     variation       119250
                     /note="2173I4: a; 114N19: g"
                     /replace="g"
     variation       119718
                     /note="2173I4: g; 114N19: t"
                     /replace="t"
     variation       120914
                     /note="2173I4: c; 114N19: a"
                     /replace="a"
     variation       121395. .121413
                     /note="2173I4: caaaaaaaaaaaaaaaaa; 114N19: c"
                     /replace="c"
     variation       123019
                     /note="2173I4: g; 114N19: a"
                     /replace="a"
     variation       123721
                     /note="2173I4: c; 114N19: t"
                     /replace="t"
     variation       123900
                     /note="2173I4: t; 114N19: c"
                     /replace="c"
     variation       123926. .123927
                     /note="2173I4: tt; 114N19: t"
                     /replace="t"
     variation       124015
                     /note="2173I4: t; 114N19: c"
                     /replace="c"
     variation       124182
                     /note="2173I4: g; 114N19: a"
                     /replace="a"
BASE COUNT    43520 a  24500 c  22595 g  33732 t
ORIGIN
```

```
  Query Match             5.9%;  Score 58.8;  DB 9;  Length 124347;
  Best Local Similarity   61.8%;  Pred. No. 0.0018;
  Matches  110;  Conservative    0;  Mismatches   67;  Indels    1;  Gaps    1;

Qy        814 GGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCA 873
              | |  |||||   | ||||||||||| |||||  ||| |  |||||| ||||||    |
Db      41340 GAAATGGGAAGATGCAGGTCAGAGGATACAAAGTAGCTGATATGTAGGATGAACAAGTCT 41399

Qy        874 GAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTT 933
              | |        | | ||   |      || |     |  |  |  ||| |||  ||
Db      41400 AGAGATCTAAAGTATAACATGAGAACTACAGATAATAAAACTGTACT-ATATATGGGATT 41458

Qy        934 CATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 991
              |||||  || | |||   |||||||| ||||||||||||  |||| ||||||||||||
Db      41459 CATGCTAAGTAAGTACATTTTAGCTGTTCTTGCCACAAACCAAAACAAAAAAAAAAAA 41516


RESULT 30
AC093574/c
LOCUS       AC093574              151993 bp    DNA     linear   PRI 23-FEB-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-178K11, complete sequence.
ACCESSION   AC093574 AC021597
VERSION     AC093574.2  GI:18873861
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151993)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151993)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 151993)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Feb 23, 2002 this sequence version replaced gi:15431100.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-178K11 (bc0286)
            ------------- Summary Statistics
                Sequencing vector: unknown; 56% of reads
                Sequencing vector: plasmid; 44% of reads
                Chemistry: Dye-terminator ET; 49% of reads
                Chemistry: Dye-terminator Big Dye; 51% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 151851 bases at least Q40
                Consensus quality: 151973 bases at least Q30
                Consensus quality: 151993 bases at least Q20
                Insert size: 158451; sum-of-contigs
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5': RP11-56B20 (UWGC:bc0174) AC026673
                3': Mapping in progress
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
```

estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8949	4503	4596	6422	6405
6	<800	6382	6584	2067	2089
6912	6827	512	<800	6249	6405
2142	2138	449	<800	470	<800
1633	1601	2282	2301	3028	2892
185	<800	404	<800	10089	9902
33	<800	3748	3800	2765	2775
235	<800	882	868	1438	1439
986	959	496	<800	5480	5750
5804	5786	1453	1429	8688	8596
166	<800	207	<800	2063	2089
2553	2584	4653	4596	7872	7784
73	<800	5302	5418	2057	2089
8410	8322	2615	2732	1970	2089
1588	1601	423	<800	344	<800
1392	1382	156	<800	5861	5750
3402	3455	3436	3385	965	988
147	<800	601	<800	4108	3983
1254	1250	52	<800	3172	3167
973	959	2019	2022	1145	1094
161	<800	3795	3800	1593	1595
313	<800	1540	1503	5146	5297
3713	3718	10075	9954	3644	3587
2627	2584	5447	5418	2365	2381
2804	2824	3913	3800	6608	6405
782	806	2683	2867	3181	3167
305	<800	622	<800	1074	988
2256	2264	1811	1749	1702	1736
4178	4181	1519	1503	759	<800
964	959	2068	2022	9566	9902
1486	1472	3072	3098	1419	1439
3586	3455	84	<800	9853	9902
6519	6434	381	<800	783	788
1447	1472	9	<800	4739	4629
2318	2264	534	<800	238	<800
1880	1840	388	<800	529	<800
1241	1250	2365	2540	2875	2892
3064	3099	5565	5418	205	<800
3415	3455	5400	5418	1223	1144
4194	4181	8055	8069	2821	2892
2586	2584	2087	2022	238	<800
2823	2824	857	868	939	988
3419	3455	613	<800	3195	3167
1501	1472	837	868	5065	4986
20	<800	3075	3216	1237	1247
4338	4390	6244	6292	5864	5750
5735	5786	3337	3385	1092	1094
4197	4181	2253	2301	379	<800
10513	10499	2916	2867	5118	4986
907	959	277	<800	989	988
842	806	1915	1912		
905	959	541	<800		
78	<800	574	<800		
12446	12330	7293	7311		
1636	1601	9275	9237		
1342	1382	2546	2622		
2807	2824	806	868		

```
  Query Match             5.9%;  Score 58.6;  DB 9;  Length 151993;
  Best Local Similarity   61.0%;  Pred. No. 0.0021;
  Matches  114;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    1;

Qy        797 GGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCAT 856
              |||  | ||| ||    | |  | | |   | |||||||| |  |||||  |||   ||
Db      36052 GGATGTGGGGAGGGCAGGAAATGAGGAGATGTAGGTCAGAGAATACAAAGTGGCAAATAT 35993

Qy        857 GTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCAC 916
              |||  || ||| |  ||||      |       ||| | |  |||  || |   || ||
Db      35992 GTATGATTAACAGGTCAGATCTAATATACAACATGAGAATTATAGGTAATAAAATTGTAC 35933

Qy        917 TGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAA 976
              |    |||| ||   ||||||| ||||||||   ||||| ||||||||||||| |||| |
Db      35932 T----GTATGTGAGATTCATGCTAAATGAGTAGATTTTACCTGCTCTTGCCACGAAAACA 35877

Qy        977 AAAAAAA 983
              || | ||
Db      35876 AACAGAA 35870
```

RESULT 31
AC012087
LOCUS AC012087 194635 bp DNA linear PRI 18-AUG-2000
DEFINITION Homo sapiens 3 BAC RP11-317L10,RP11-150K6,RP11-68D12 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC012087 AC011608 AC012034 AC012016
VERSION AC012087.10 GI:7114465
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194635)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K., Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,D.K., Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,J., Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194635)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 194635)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 194635)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On or before Aug 18, 2000 this sequence version replaced gi:6067112, gi:7025646, gi:7008674, gi:7025648.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

---------- Summary Statistics ----------

Contig length:	194635
Phrap values in estimate:	193865
Average error rate (BCM-Phrap estimate):	3.31225e-05
Fraction of Phrap values less than 40 :	0.00865035
Number of consensus changing edits:	46
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
4981	agatggagat(n)tgatgaggtg	agatggagat(g)tgatgaggtg
18700	agaactttta(n)ccacccagcc	agaactttta(a)ccacccagcc
23735	tctaggagcc(n)taatagaagt	tctaggagcc(c)taatagaagt
44381	ctgaatttga(n)tttcatataa	ctgaatttga(a)tttcatataa
53445	aaaaaaaaaa(n)tcattgccaa	aaaaaaaaaa(a)tcattgccaa
61587	tgcaaaaaag(c)agatgggatc	tgcaaaaaag(t)agatgggatc
89906	ttcaatggtg(g)tattgtgagt	ttcaatggtg(a)tattgtgagt
97122	caggcgccct(c)ccccgctct	caggcgccct(a)ccccgctct
97318	agcccctggc(c)gtgccccggg	agcccctggc(t)gtgccccggg
97464	tgctctgaag(g)tccctgagcc	tgctctgaag(c)tccctgagcc
97608	aagaacaggg(a)gggtggggcc	aagaacaggg(c)gggtggggcc
98041	gcttttgcct(n)tggttgaaga	gcttttgcct(a)tggttgaaga
101579	caatctgctg(n)agtatgttgt	caatctgctg(t)agtatgttgt
106238	aagcagtatg(n)cttcatcttt	aagcagtatg(a)cttcatcttt
106254	tctttttttg(n)ttttttttt	tctttttttg(t)ttttttttt
108707	ataagcatgg(c)cttttgactt	ataagcatgg(t)cttttgactt
114218	gtaaaatttg(g)taaaatgaaa	gtaaaatttg(a)taaaatgaaa
123681	gcgatggctg(c)ngnngnnggc	gcgatggctg(g)cggcggcggc
123682	cgatggctgc(n)gnngnnggcg	cgatggctgg(c)ggcggcggcg

```
123684     atggctgcng(n)ngnnggcggg     atggctggcg(g)cggcggcggg
123685     tggctgcngn(n)gnnggcgggg     tggctggcgg(c)ggcggcgggg
123687     gctgcngnng(n)nggcggggtg     gctggcggcg(g)cggcggggtg
123688     ctgcngnngn(n)ggcggggtgt     ctggcggcgg(c)ggcggggtgt
123702     ggggtgtgcg(g)caggaggcca     ggggtgtgcg(c)caggaggcca
123720     ccatttcctc(c)ctgcgcccct     ccatttcctc(g)ctgcgcccct
123731     ctgcgcccct(n)gcggagccgg     ctgcgcccct(g)gcggagccgg
123763     cttggccgcc(n)ccgccaccgc     cttggccgcc(g)ccgccaccgc
130680     taggaaagga(n)tggtctggac     taggaaagga(t)tggtctggac
137763     aaaaaaaaaa(n)ccagaaagaa     aaaaaaaaaa(a)acagaaagaa
137764     aaaaaaaaan(c)cagaaagaaa     aaaaaaaaaa(a)cagaaagaaa
148155     gaggtcagga(n)ttcgagacca     gaggtcagga(a)ttcgagacca
151494     ttatctaaga(c)atacatttct     ttatctaaga(a)atacatttct
163884     aaaaaaaaaa(n)aaagtaagtc     aaaaaaaaaa(a)aaagtaagtc
171028     tacttcataa(n)taattaccaa     tacttcataa(a)taattaccaa
181695     tgagaacatg(c)aatatttgtc     tgagaacatg(g)aatatttgtc
185040     agctggaaca(n)tcaagcctct     agctggaaca(a)tcaagcctct
185127     gtgaattcct(n)actnccatcc     gtgaattcct(t)actcccatcc
185131     attcctnact(n)ccatccaggg     attccttact(c)ccatccaggg
185153     cttttcccca(n)gtcattaatg     cttttcccca(t)gtcattaatg
185191     caccttatga(n)gnaaataann     caccttatga(g)ggaaataact
185193     ccttatgang(n)aaataannga     ccttatgagg(g)aaataactga
185200     angnaaataa(n)ngagggcagt     agggaaataa(c)tgagggcagt
185201     ngnaaataan(n)gagggcagtc     gggaaataac(t)gagggcagtc
190487     acaaacagca(n)cactgctggt     acaaacagca(a)cactgctggt
190499     actgctggta(n)cactggcaag     actgctggta(a)cactggcaag
192055     aagcctgttg(n)tctagacagt     aagcctgttg(t)tctagacagt
```

```
--------- Distribution of Quality < 40 Bases --------------------

        1000|
         900|
         800|
   #     700|
  bases  600|                                  *    *
         500|                                  *    *
         400|                                  *    *
         300|                                  *    *
         200|                                  *    *
         100|                        *    *    *    *
           0|        *    *    *     *    *    *    *
            ------------------------------------------
              5   10   15   20   25   30   35   40
                     Phrap Value Range
```

```
--------------------------------------------------------------------
           Version: 1.01   qxfo.
FEATURES             Location/Qualifiers
     source          1. .194635
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-317L10,RP11-150K6,RP11-68D12"
     repeat_region   complement(40. .496)
                     /rpt_family="L1M4"
     repeat_region   1423. .1452
                     /rpt_family="AT_rich"
     repeat_region   1705. .2095
                     /rpt_family="LTR16A"
     repeat_region   complement(2399. .2430)
                     /rpt_family="L2"
     repeat_region   complement(2431. .2954)
                     /rpt_family="MLT1F"
     repeat_region   complement(2955. .3056)
                     /rpt_family="L2"
     repeat_region   complement(3109. .3199)
                     /rpt_family="L2"
     repeat_region   5566. .5781
                     /rpt_family="MER20"
     repeat_region   6399. .6418
                     /rpt_family="(T)n"
     repeat_region   complement(6497. .6702)
                     /rpt_family="L1ME3A"
     repeat_region   7342. .7405
                     /rpt_family="AT_rich"

  Query Match              5.9%;  Score 58.6;  DB 9;  Length 194635;
  Best Local Similarity   61.0%;  Pred. No. 0.002;
  Matches  114;  Conservative    0;  Mismatches   69;  Indels    4;  Gaps    1;

Qy        797 GGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCAT 856
              |||  |  ||| ||       | |  |  | |     | |||||||||  |  ||||  |||   ||
Db      54877 GGATGTGGGGAGGGCAGGAAATGAGGAGATGTAGGTCAGAGAATACAAAGTGGCAAATAT 54936

Qy        857 GTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCAC 916
              |||  || ||| |  ||||        |          ||| |  |  |||  || |   || ||
Db      54937 GTATGATTAACAGGTCAGATCTAATATACAACATGAGAATTATAGGTAATAAAATTGTAC 54996

Qy        917 TGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAA 976
              |    |||| ||    ||||||| ||||||||     ||||| |||||||||||||  |||| |
Db      54997 T----GTATGTGAGATTCATGCTAAATGAGTAGATTTTACCTGCTCTTGCCACGAAAACA 55052

Qy        977 AAAAAAA 983
              ||  |  ||
Db      55053 AACAGAA 55059
```

```
RESULT 32
HSM130AC2/c
LOCUS       HSM130AC2               4950 bp    mRNA    linear   PRI 22-NOV-1996
DEFINITION  H.sapiens mRNA for M130 antigen cytoplasmic variant 2.
ACCESSION   Z22970
VERSION     Z22970.1  GI:312145
KEYWORDS    antigen; antigen M130.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4950)
  AUTHORS   Law,S.K., Micklem,K.J., Shaw,J.M., Zhang,X.P., Dong,Y., Willis,A.C.
            and Mason,D.Y.
  TITLE     A new macrophage differentiation antigen which is a member of the
            scavenger receptor superfamily
  JOURNAL   Eur. J. Immunol. 23 (9), 2320-2325 (1993)
  MEDLINE   93380506
   PUBMED   8370408
REFERENCE   2  (bases 1 to 4950)
  AUTHORS   Micklem,K.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-1993) Kingsley K.J. Micklem, Nuffield Department
            of Pathology, University of Oxford, Level 1, Maternity Block, John
            Radcliffe, Hospital, Headington, Oxford, OX3 9DU, United Kingdom
FEATURES             Location/Qualifiers
     source          1. .4950
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="Macrophage"
     CDS             102. .3572
                     /codon_start=1
                     /product="M130 antigen cytoplasmic variant 2"
                     /protein_id="CAA80543.1"
                     /db_xref="GI:312146"
                     /db_xref="SPTREMBL:Q07900"
                     /translation="MVLLEDSGSADFRRHFVNLSPFTITVVLLLSACFVTSSLGGTDK
                     ELRLVDGENKCSGRVEVKVQEEWGTVCNNGWSMEAVSVICNQLGCPTAIKAPGWANSS
                     AGSGRIWMDHVSCRGNESALWDCKHDGWGKHSNCTHQQDAGVTCSDGSNLEMRLTRGG
                     NMCSGRIEIKFQGRWGTVCDDNFNIDHASVICRQLECGSAVSFSGSSNFGEGSGPIWF
                     DDLICNGNESALWNCKHQGWGKHNCDHAEDAGVICSKGADLSLRLVDGVTECSGRLEV
                     RFQGEWGTICDDGWDSYDAAVACKQLGCPTAVTAIGRVNASKGFGHIWLDSVSCQGHE
                     PAVWQCKHHEWGKHYCNHNEDAGVTCSDGSDLELRLRGGGSRCAGTVEVEIQRLLGKV
                     CDRGWGLKEADVVCRQLGCGSALKTSYQVYSKIQATNTWLFLSSCNGNETSLWDCKNW
                     QWGGLTCDHYEEAKITCSAHREPRLVGGDIPCSGRVEVKHGDTWGSICDSDFSLEAAS
```

```
                     VLCRELQCGTVVSILGGAHFGEGNGQIWAEEFQCEGHESHLSLCPVAPRPEGTCSHSR
                     DVGVVCSRYTEIRLVNGKTPCEGRVELKTLGAWGSLCNSHWDIEDAHVLCQQLKCGVA
                     LSTPGGARFGKGNGQIWRHMFHCTGTEQHMGDCPVTALGASLCPSEQVASVICSGNQS
                     QTLSSCNSSSLGPTRPTIPEESAVACIESGQLRLVNGGGRCAGRVEIYHEGSWGTICD
                     DSWDLSDAHVVCRQLGCGEAINATGSAHFGEGTGPIWLDEMKCNGKESRIWQCHSHGW
                     GQQNCRHKEDAGVICSEFMSLRLTSEASREACAGRLEVFYNGAWGTVGKSSMSETTVG
                     VVCRQLGCADKGKINPASLDKAMSIPMWVDNVQCPKGPDTLWQCPSSPWEKRLASPSE
                     ETWITCDNKIRLQEGPTSCSGRVEIWHGGSWGTVCDDSWDLDDAQVVCQQLGCGPALK
                     AFKEAEFGQGTGPIWLNEVKCKGNESSLWDCPARRWGHSECGHKEDAAVNCTDISVQK
                     TPQKATTGRSSRQSSFIAVGILGVVLLAIFVALFFLTKKRRQRQRLAVSSRGENLVHQ
                     IQYREMNSCLNADDLDLMNSSGLWVLGGSIAQGFRSVAAVEAQTFYFDKQLKKSKNVI
                     GSLDAYNGQE"
     sig_peptide     102. .221
     mat_peptide     222. .3569
                     /product="M130 antigen cytoplasmic variant 2"
                     /evidence=experimental
     polyA_signal    4924. .4929
     polyA_signal    4928. .4933
     polyA_signal    4932. .4937
     polyA_site      4947
BASE COUNT     1348 a    970 c   1277 g   1355 t
ORIGIN

  Query Match             5.9%;  Score 58.4;  DB 9;  Length 4950;
  Best Local Similarity   59.0%;  Pred. No. 0.0025;
  Matches  118;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        787 TGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAG 846
              ||| | | | | | ||| | | | | | | | |||||||||||  |||||
Db       4080 TGACAGGAAATGGGAATGGAAGAAACTGTGAAAGACGTATGTAGGTCAGAGGATACAAAG 4021

Qy        847 CAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCAT 906
              |||||  | ||||||||||||  |  | | |  | | ||  |
Db       4020 TAGCAGATACGTAGAATGAACCAGTCTAGAGGTCTAATGTACAACATGAGGATTATAGGT 3961

Qy        907 CAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGC 966
              |  || | |||| | |  ||||||| || |||||  ||||||||||||||||||
Db       3960 AATAACATTGTGCT-GTATGTAGGATTCATGCTAAGTGAGTAGGTTTTAGCTGCTCTTGC 3902

Qy        967 CACAAAAAAAAAAAAAAAAA 986
              ||||| || || | ||||||
Db       3901 CACAAGAACAACACAAAAAA 3882


RESULT 33
AC131206
LOCUS       AC131206               177260 bp    DNA     linear   PRI 31-DEC-2002
DEFINITION  Homo sapiens 12 BAC RP11-157G21 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC131206
VERSION     AC131206.2  GI:22549585
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177260)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177260)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 177260)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 177260)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 177260)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:22296911.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
```

continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .177260
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-157G21"
     misc_feature    1. .2001
                     /note="overlaps bases 44655. .46655 of clone AC131207"
                     /function="clone overlap"
     STS             64444. .64708
                     /standard_name="D12S1890"
     STS             68193. .68303
                     /standard_name="D10S275"
     STS             78117. .78366
                     /standard_name="STS-H93331"
     STS             90052. .90229
                     /standard_name="D11S2442"
     STS             90092. .90372
                     /standard_name="D11S3206"
     STS             93532. .93750
                     /standard_name="D11S3059"
     STS             137355. .137635
                     /standard_name="D11S3206"
     STS             147420. .147700
                     /standard_name="D11S3206"
     STS             147563. .147740
                     /standard_name="D11S2442"
     misc_feature    175236. .177260
                     /note="overlaps bases 1. .2025 of clone AC006927"
                     /function="clone overlap"
BASE COUNT    59476 a  33910 c  33289 g  50585 t
ORIGIN

  Query Match             5.9%;  Score 58.4;  DB 9;  Length 177260;
  Best Local Similarity   59.0%;  Pred. No. 0.0023;
  Matches  118;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        787 TGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAG 846
              ||| | | | | |  |||  |   | |   | |   |  | ||||||||||||  |||||
Db      64219 TGACAGGAAATGGGAATGGAAGAAACTGTGAAAGACGTATGTAGGTCAGAGGATACAAAG 64278

Qy        847 CAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCAT 906
              |||||  | ||||||||||||   |     | | |       | | |||          |
Db      64279 TAGCAGATACGTAGAATGAACCAGTCTAGAGGTCTAATGTACAACATGAGGATTATAGGT 64338

Qy        907 CAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGC 966
              |        || | |||| | |  ||||||| || |||||  ||||||||||||||||||
Db      64339 AATAACATTGTGCT-GTATGTAGGATTCATGCTAAGTGAGTAGGTTTTAGCTGCTCTTGC 64397

Qy        967 CACAAAAAAAAAAAAAAAAA 986
              ||||| || || | ||||||
Db      64398 CACAAGAACAACACAAAAAA 64417

RESULT 34
```

```
AC023003
LOCUS       AC023003               140055 bp    DNA     linear   HTG 01-MAR-2000
DEFINITION  Homo sapiens clone RP11-270M20, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION   AC023003
VERSION     AC023003.2  GI:7139812
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140055)
  AUTHORS   Birren,B., Linton,L. and Lander,E.
  TITLE     Homo sapiens, clone RP11-270M20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140055)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 1, 2000 this sequence version replaced gi:6921586.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6597
                Center clone name: 270_M_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 133511 bases at least Q40
                Consensus quality: 136178 bases at least Q30
                Consensus quality: 137450 bases at least Q20
                Insert size: 139000; agarose-fp
                Insert size: 138955; sum-of-contigs
                Quality coverage: 5.1 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1060: contig of 1060 bp in length
            *     1061     1160: gap of 100 bp
            *     1161     3767: contig of 2607 bp in length
            *     3768     3867: gap of 100 bp
            *     3868     8490: contig of 4623 bp in length
```

```
            *     8491     8590: gap of 100 bp
            *     8591    14297: contig of 5707 bp in length
            *    14298    14397: gap of 100 bp
            *    14398    20798: contig of 6401 bp in length
            *    20799    20898: gap of 100 bp
            *    20899    31912: contig of 11014 bp in length
            *    31913    32012: gap of 100 bp
            *    32013    43544: contig of 11532 bp in length
            *    43545    43644: gap of 100 bp
            *    43645    52029: contig of 8385 bp in length
            *    52030    52129: gap of 100 bp
            *    52130    67985: contig of 15856 bp in length
            *    67986    68085: gap of 100 bp
            *    68086    84647: contig of 16562 bp in length
            *    84648    84747: gap of 100 bp
            *    84748   108247: contig of 23500 bp in length
            *   108248   108347: gap of 100 bp
            *   108348   140055: contig of 31708 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140055
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-270M20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1060
                     /note="assembly_fragment"
     misc_feature    1161. .3767
                     /note="assembly_fragment"
     misc_feature    3868. .8490
                     /note="assembly_fragment"
     misc_feature    8591. .14297
                     /note="assembly_fragment"
     misc_feature    14398. .20798
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    20899. .31912
                     /note="assembly_fragment"
     misc_feature    32013. .43544
                     /note="assembly_fragment"
     misc_feature    43645. .52029
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    52130. .67985
                     /note="assembly_fragment"
     misc_feature    68086. .84647
                     /note="assembly_fragment"
     misc_feature    84748. .108247
                     /note="assembly_fragment"
     misc_feature    108348. .140055
                     /note="assembly_fragment"
BASE COUNT    42310 a  26393 c  26939 g  43312 t   1101 others
ORIGIN

  Query Match             5.8%;  Score 58;  DB 2;  Length 140055;
  Best Local Similarity   61.9%;  Pred. No. 0.0029;
  Matches  125;  Conservative    0;  Mismatches   75;  Indels    2;  Gaps    2;

Qy        787 TGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAG 846
              ||| || ||||          |  |  | |  ||||||     |||||||||||| |||||
Db      15860 TGACAGAGGAGGGTAGAAAATGCATAGGAAATGGGAATATATAGGTCAGAGGACACAAAG 15919

Qy        847 CAGCAGCCATGTAGAATGAACC-GTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCA 905
              |||||||||||  ||  | ||  |  |||  |||||  |  |||          |    ||  |
Db      15920 CAGCAGCCATGCAGGAGGACCAAGTCTAGAGATCTAAGGTACATTATGAGGACCATACAG 15979

Qy        906 TCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTG 965
                |   |  |      ||  |  |||  |||||||  |||  |||  |  |||  ||||||||||
Db      15980 GTAATAAGAGCGTACTGTGTATGGGATTCATGCTAAACTAGTA-GATTTGGCTGCTCTTG 16038

Qy        966 CCACAAAAAAAAAAAAAAAAA 987
              |||||||||  |||  ||||||||
Db      16039 CCACAAAAACAAACAAAAAAA 16060


RESULT 35
AL139342
LOCUS       AL139342              151553 bp    DNA     linear   PRI 03-JAN-2002
DEFINITION  Human DNA sequence from clone RP5-1016N21 on chromosome 1q42.13-43.
            Contains part of the gene for a novel protein (ortholog of
            Drosophila Pecanex (PCX), similar to KIAA0805), part of the gene
            for a novel protein similar to KIAA0995, STSs and GSSs, complete
            sequence.
ACCESSION   AL139342
VERSION     AL139342.7  GI:9796485
KEYWORDS    HTG; KIAA0805; KIAA0995; PCX; Pecanex.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151553)
  AUTHORS   Cobley,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9712646.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            This sequence is the entire insert of clone RP5-1016N21 The true
            left end of clone RP11-109B17 is at 90706 in this sequence. The
            true right end of clone RP11-109B17 is at 77343 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP5-1016N21 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .151553
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q42.13-43"
                     /clone="RP5-1016N21"
                     /clone_lib="RPCI-5"
     repeat_region   1. .77
                     /note="Alu repeat: matches 232. .310 of consensus"
     repeat_region   101. .408
                     /note="L1MC/D repeat: matches 5263. .5571 of consensus"
     repeat_region   465. .619
                     /note="L1MD2 repeat: matches 5810. .5953 of consensus"
     repeat_region   628. .729
                     /note="L1MEc repeat: matches 2110. .2213 of consensus"
     repeat_region   836. .1468
                     /note="L1M4 repeat: matches 4267. .4889 of consensus"
     repeat_region   1511. .1685
```

```
                     /note="L1MC3 repeat: matches 6433. .6606 of consensus"
repeat_region        1745. .2410
                     /note="L1MC3 repeat: matches 6609. .7281 of consensus"
repeat_region        2411. .2705
                     /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region        2706. .3163
                     /note="L1MC3 repeat: matches 7281. .7736 of consensus"
repeat_region        3183. .3479
                     /note="AluSg repeat: matches 1. .294 of consensus"
misc_feature         complement(4350. .4820)
                     /note="match: GSS: Em:AQ827082"
repeat_region        4999. .5225
                     /note="L2 repeat: matches 1328. .1564 of consensus"
repeat_region        5811. .6105
                     /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region        6180. .6371
                     /note="L1MC3 repeat: matches 5725. .5926 of consensus"
repeat_region        6386. .6421
                     /note="18 copies 2 mer gt 97% conserved"
repeat_region        6673. .7475
                     /note="L2 repeat: matches 1777. .2614 of consensus"
repeat_region        7485. .7709
                     /note="MIR repeat: matches 21. .262 of consensus"
repeat_region        7733. .7783
                     /note="L2 repeat: matches 2648. .2705 of consensus"
repeat_region        8206. .8517
                     /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region        8833. .10173
                     /note="pTR5 repeat: matches 743. .2438 of consensus"
repeat_region        10349. .10386
                     /note="19 copies 2 mer tg 100% conserved"
repeat_region        10388. .10664
                     /note="AluY repeat: matches 2. .279 of consensus"
repeat_region        11826. .12145
                     /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region        12656. .12756
                     /note="MIR repeat: matches 50. .157 of consensus"
repeat_region        12776. .13081
                     /note="AluSq repeat: matches 1. .306 of consensus"
misc_feature         complement(13145. .13540)
                     /note="match: GSS: Em:AQ177321"
repeat_region        13457. .13718
                     /note="AluJo repeat: matches 1. .258 of consensus"
repeat_region        14443. .14562
                     /note="6 copies 20 mer 78% conserved"
repeat_region        14515. .14558
                     /note="22 copies 2 mer ag 77% conserved"
gene                 complement(join(14896. .15073,19597. .19738,40166. .40316,
                     41154. .41246,57685. .57781,58949. .59058,78822. .78913,
                     80037. .80147,88401. .88572,97782. .97814,97914. .98067,
                     107109. .107254))
                     /gene="dJ1016N21.1"
CDS                  complement(join(<14896. .15073,19597. .19738,40166. .40316,
                     41154. .41246,57685. .57781,58949. .59058,78822. .78913,
                     80037. .80147,88401. .88572,97782. .97814,97914. .98067,
                     107109. .>107254))
                     /gene="dJ1016N21.1"
                     /note="match: cDNAs: Em:AF096286 Em:AB018348
                     match: proteins: Tr:Q9QYC1 Tr:O94897 Tr:Q9W507 Tr:O46074
                     Tr:Q9W506 Tr:O61825 Sw:P18490"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ1016N21.1 (novel protein (ortholog of
                     Drosophila Pecanex (PCX), similar to KIAA0805))"
                     /protein_id="CAD20144.1"
                     /db_xref="GI:18070856"
                     /translation="HVSQDLEASSCSSTQGKFNREQFYKFIIFPPGKWIKVWYDRLTLL
                     ALLDRTEDIKENVLAILLIVLVSLLGFLTLSQGFCKDMWVLLFCLVMASCQYSLLKSV
                     QPDPASPIHGHNQIITYSRPIYFCVLCGLILLLDTGAKARHPPSYVVYGLKLFSPVFL
                     QSARDYLIVFLYCFPAISLLGLFPQINTFCTYLLEQIDMLFFGGSAVSGITSAVYSVA
                     RSVLAAALLHAVCFSAVKEPWSMQHIPALFSAFCGLLVALSYHLSRQSSDPSVLMSFI
                     QCRLFPKFLHQNLAESAADPLPKKMKDSVTDVLKWDLIVCAVVAVLSFAVSASTVFLS
                     LRPFLSIVLFALAGAVGFVTHYVLPQLRKHHPWMWISHPILKNKEYHQREVRDVAHLM
                     WFERLYVWLQCFEKYILYPALILNALTIDAFLISNHRRLGTHWDIFLMIIAGMKLLRT
                     SFCNPVYQFINLSFTVIFFHFDYKDISESFLLDFFMVSILFSK"
repeat_region        15184. .15344
                     /note="MIR repeat: matches 12. .179 of consensus"
repeat_region        16957. .17561
                     /note="MER6A repeat: matches 1. .598 of consensus"
misc_feature         complement(18841. .19304)
                     /gene="dJ1016N21.1"
                     /note="match: GSS: Em:AQ558297"
misc_feature         complement(18915. .19311)
                     /gene="dJ1016N21.1"
                     /note="match: GSS: Em:AQ338718"
repeat_region        20315. .20425
                     /note="MIR repeat: matches 18. .129 of consensus"
repeat_region        21268. .21570
                     /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region        21571. .21690
                     /note="6 copies 20 mer 80% conserved"
repeat_region        22427. .23076
                     /note="L2 repeat: matches 1148. .1840 of consensus"
repeat_region        23241. .23347
                     /note="MER5A repeat: matches 4. .112 of consensus"
repeat_region        23351. .23520
                     /note="MER5A repeat: matches 26. .188 of consensus"
repeat_region        23676. .23858
                     /note="L2 repeat: matches 953. .1144 of consensus"
repeat_region        23868. .24080
                     /note="AluJb repeat: matches 87. .298 of consensus"
repeat_region        24270. .24836
                     /note="L1ME repeat: matches 5092. .5651 of consensus"
repeat_region        24837. .25148
                     /note="AluY repeat: matches 1. .310 of consensus"
misc_feature         complement(24980. .25609)
                     /gene="dJ1016N21.1"
                     /note="match: GSS: Em:AQ344097"
repeat_region        25149. .25356
                     /note="L1ME repeat: matches 4886. .5093 of consensus"
repeat_region        25566. .25747
                     /note="MER97b repeat: matches 1. .187 of consensus"
misc_feature         complement(25605. .26037)
                     /gene="dJ1016N21.1"
                     /note="match: GSS: Em:AQ683895"
repeat_region        25810. .25859
                     /note="Tigger4(Zombi) repeat: matches -1. .48 of consensus"
repeat_region        25860. .25907
                     /note="MER46A repeat: matches 49. .97 of consensus"
repeat_region        25900. .25979
                     /note="MER46A repeat: matches 153. .236 of consensus"
misc_feature         26937. .27418
                     /note="match: GSS: Em:AQ561984"
repeat_region        27022. .27459
                     /note="TIGGER2 repeat: matches 2280. .2716 of consensus"
repeat_region        27471. .27530
                     /note="MER8 repeat: matches 178. .239 of consensus"
repeat_region        28304. .28584
                     /note="AluSg repeat: matches 1. .310 of consensus"
misc_feature         29471. .29912
                     /note="match: GSS: Em:B66297"
misc_feature         29498. .29966
                     /note="match: GSS: Em:AQ229439"
```

```
Query Match             5.8%;  Score 58;  DB 9;  Length 151553;
Best Local Similarity   59.7%;  Pred. No. 0.0029;
Matches  117;  Conservative    0;  Mismatches   75;  Indels    4;  Gaps    1;

Qy        789 AGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCA 848
              ||| || | | |  ||||||    ||    ||| ||  |   || |||       |||| |
Db      47275 AGTTAACATGGGCATGGGGGAGGTAGAGATGGGGATATGTGTGTTAGAAATTATAAAGTA 47334

Qy        849 GCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCA 908
              |||  |||||| || |||    |       | |         ||| |||  |    || |
Db      47335 GCAGATATGTAGGATAAACAACTCCAGCGATCTAATGTACATGAGAACTATAATTAATAA 47394
```

```
Qy        909 GCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCA 968
              || |||     |||||||  ||| ||||||||| |||   |||||||||||| ||||||
Db      47395 AACTGTACT----ATATTTGGGATTCCTGCAAAATAAGTAGATTTTAGCTGCTTTTGCCA 47450

Qy        969 CAAAAAAAAAAAAAAA 984
              |||||| ||| |||||
Db      47451 CAAAAACAAATAAAAA 47466
```

RESULT 36
AC019159/c
LOCUS AC019159 163085 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-56O18 from 2, complete sequence.
ACCESSION AC019159
VERSION AC019159.8 GI:13677116
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163085)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 163085)
AUTHORS Goyea,E., Cotton,M., Spalding,L. and Lehnert,L.
TITLE The sequence of Homo sapiens BAC clone RP11-56O18
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 163085)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 163085)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 163085)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 163085)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Apr 19, 2001 this sequence version replaced gi:11276269.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0056O18
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC027111; the clone sequenced to the right is AC023003. Actual start of this clone is at base position 1 of RP11-56O18; actual end is at base position 163085 of RP11-56O18.

The sequence from base position 822245 to 82358 is derived from a single plasmid subclone. Size information based on PCR supports the assembly.

FEATURES Location/Qualifiers
source 1. .163085
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-56O18"
/clone_lib="RPCI-11"
misc_feature 1. .51
/note="similar to EST AV696391 (NID:g10298254)"
misc_feature 107. .224
/note="similar to EST AV696391 (NID:g10298254)"
repeat_region 197. .373
/rpt_family="MIR"
repeat_region 568. .667
/rpt_family="MIR"
repeat_region 668. .775
/rpt_family="MER1_type"
repeat_region 1249. .1439
/rpt_family="MIR"
repeat_region 8077. .8445
/rpt_family="MaLR"
repeat_region 8451. .9804
/rpt_family="MaLR"
repeat_region 9861. .10142
/rpt_family="Alu"
repeat_region 10149. .10387
/rpt_family="MaLR"
repeat_region 10393. .10790
/rpt_family="MaLR"
repeat_region 11489. .11710
/rpt_family="MIR"
repeat_region 11985. .12022
/rpt_family="(TG)n"
repeat_region 12131. .12886
/rpt_family="L1"
repeat_region 13153. .13392
/rpt_family="L1"
repeat_region 13397. .13441
/rpt_family="(TG)n"
repeat_region 13513. .13577
/rpt_family="(TA)n"
repeat_region 13578. .14029
/rpt_family="L1"

```
     repeat_region   14082. .14306
                     /rpt_family="L1"
     repeat_region   14739. .14778
                     /rpt_family="L2"
     repeat_region   15455. .15476
                     /rpt_family="AT_rich"
     repeat_region   16189. .16500
                     /rpt_family="Alu"
     repeat_region   16601. .16736
                     /rpt_family="MIR"
     repeat_region   19100. .19209
                     /rpt_family="MIR"
     repeat_region   19257. .19388
                     /rpt_family="MIR"
     repeat_region   19840. .19906
                     /rpt_family="(TGAA)n"
     repeat_region   20239. .20271
                     /rpt_family="(CA)n"
     repeat_region   20643. .20923
                     /rpt_family="Alu"
     repeat_region   21036. .21437
                     /rpt_family="MaLR"
     repeat_region   21463. .21653
                     /rpt_family="MIR"
     repeat_region   21736. .21861
                     /rpt_family="MIR"
     repeat_region   22171. .22330
                     /rpt_family="Alu"
     repeat_region   22695. .22809
                     /rpt_family="MIR"
     repeat_region   23975. .24015
                     /rpt_family="A-rich"
     repeat_region   24958. .25194
                     /rpt_family="MIR"
     repeat_region   25352. .25425
                     /rpt_family="L1"
     repeat_region   26112. .26417
                     /rpt_family="L1"
     repeat_region   26739. .26949
                     /rpt_family="MIR"
     repeat_region   28668. .28791
                     /rpt_family="(TA)n"
     repeat_region   29118. .29628
                     /rpt_family="L2"
     repeat_region   30308. .30338
                     /rpt_family="(TTA)n"
     repeat_region   30339. .30881
                     /rpt_family="L1"
     repeat_region   30882. .30904
                     /rpt_family="(TA)n"
     repeat_region   30960. .31365
                     /rpt_family="Alu"
     repeat_region   31375. .31426
                     /rpt_family="Alu"
     repeat_region   31541. .31626
                     /rpt_family="MIR"
     repeat_region   32254. .32329
                     /rpt_family="MIR"
     repeat_region   32646. .33176
                     /rpt_family="L2"
     repeat_region   33297. .33445
                     /rpt_family="L2"
     repeat_region   33566. .33749
                     /rpt_family="MIR"
     repeat_region   34342. .34453

  Query Match             5.8%;  Score 58;  DB 9;  Length 163085;
  Best Local Similarity   61.9%;  Pred. No. 0.0029;
  Matches  125;  Conservative    0;  Mismatches   75;  Indels    2;  Gaps    2;

Qy        787 TGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAG 846
              ||| || ||||          | | | | ||||||     ||||||||||||| |||||
Db      76533 TGACAGAGGAGGGTAGAAAATGCATAGGAAATGGGAATATATAGGTCAGAGGACACAAAG 76474

Qy        847 CAGCAGCCATGTAGAATGAACC-GTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCA 905
              |||||||||||| || | || | ||| ||||| | |||         |     || | |
Db      76473 CAGCAGCCATGCAGGAGGACCAAGTCTAGAGATCTAAGGTACATTATGAGGACCATACAG 76414

Qy        906 TCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTG 965
              |    | |    || | ||| ||||||| |||  |||  | ||| ||||||||||
Db      76413 GTAATAAGAGCGTACTGTGTATGGGATTCATGCTAAACTAGTA-GATTTGGCTGCTCTTG 76355

Qy        966 CCACAAAAAAAAAAAAAAAAA 987
              ||||||||| ||| ||||||||
Db      76354 CCACAAAAACAAACAAAAAAA 76333
```

RESULT 37
AC091178
LOCUS AC091178 99268 bp DNA linear PRI 01-APR-2002
DEFINITION Homo sapiens chromosome 17, clone CTD-2206N4, complete sequence.
ACCESSION AC091178
VERSION AC091178.7 GI:19852133
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 99268)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone CTD-2206N4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 99268)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 99268)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

```
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 1, 2002 this sequence version replaced gi:18643501.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13055
                Center clone name: 2206_N_4
            --------------

            Only the first 99.2 kilobases of this clone are being submitted.
            The remainder of the clone is overlapped by either accession number
            AC006441 [WICGR project L515] or
            accession number AC004408 [WICGR project L309].
FEATURES             Location/Qualifiers
     source          1. .99268
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2206N4"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   283. .1026
                     /rpt_family="Tigger1"
     repeat_region   1032. .1324
                     /rpt_family="AluSc"
     repeat_region   1325. .1501
                     /rpt_family="Tigger1"
     repeat_region   complement(1502. .1804)
                     /rpt_family="AluSq"
     repeat_region   1805. .2245
                     /rpt_family="Tigger1"
     repeat_region   2246. .2567
                     /rpt_family="LTR28"
     repeat_region   2708. .2999
                     /rpt_family="AluSg"
     repeat_region   3369. .3396
                     /rpt_family="(TTTTTG)n"
     repeat_region   complement(3640. .3935)
                     /rpt_family="AluY"
     unsure          complement(3903. .3908)
                     /note="<30 qual SNGL region"
     unsure          complement(3915. .3919)
                     /note="<30 qual SNGL region"
     repeat_region   4083. .4237
                     /rpt_family="AluSc"
     repeat_region   complement(4938. .5073)
                     /rpt_family="AluSg"
     repeat_region   complement(5107. .5413)
                     /rpt_family="AluSq"
     repeat_region   complement(6121. .6407)
                     /rpt_family="AluSx"
     repeat_region   6553. .6870
                     /rpt_family="AluY"
     unsure          6562. .6665
                     /note="single clone coverage"
     unsure          6620. .6656
                     /note="<30 qual SNGL region"
     unsure          6733. .6741
                     /note="<30 qual SNGL region"
     unsure          6752. .6761
                     /note="<30 qual SNGL region"
     unsure          6776. .6780
                     /note="<30 qual SNGL region"
     unsure          6814. .6824
                     /note="<30 qual SNGL region"
     repeat_region   6887. .7158
                     /rpt_family="AluJo"
     unsure          7064. .7069
                     /note="<30 qual SNGL region"
     repeat_region   7161. .7227
                     /rpt_family="(GGAA)n"
     repeat_region   7385. .7474
                     /rpt_family="AluSg/x"
     repeat_region   complement(7510. .7804)
                     /rpt_family="AluSp"
     unsure          7540. .7551
                     /note="<30 qual SNGL region"
     repeat_region   9096. .9436
                     /rpt_family="AluY"
     repeat_region   13978. .14187
                     /rpt_family="MIR"
     repeat_region   15639. .15660
                     /rpt_family="AT_rich"
     repeat_region   complement(15685. .15997)
                     /rpt_family="AluJb"
     repeat_region   complement(16062. .16730)
                     /rpt_family="L1ME4A"
     repeat_region   complement(16840. .17144)
                     /rpt_family="AluSx"
     repeat_region   17854. .18011
                     /rpt_family="L1M4"
     repeat_region   complement(18099. .18424)
                     /rpt_family="MER3"
     repeat_region   18451. .18471
                     /rpt_family="AT_rich"
     repeat_region   18492. .18784
                     /rpt_family="AluJb"
     repeat_region   complement(18788. .18890)
                     /rpt_family="MER3"
     repeat_region   complement(18894. .19002)
                     /rpt_family="MER3"
     repeat_region   19029. .19140
                     /rpt_family="L1ME3"
     repeat_region   19181. .19229
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(19242. .19488)
                     /rpt_family="AluJo"
     repeat_region   complement(19497. .19849)
                     /rpt_family="AluSq"
     repeat_region   20030. .20162
                     /rpt_family="FLAM_C"
     repeat_region   20424. .20478
                     /rpt_family="(TG)n"
     repeat_region   20479. .20500
                     /rpt_family="(TGG)n"
     repeat_region   20791. .20871
                     /rpt_family="MIR3"
     repeat_region   21057. .21245
                     /rpt_family="MER58C"
     repeat_region   21249. .21539
                     /rpt_family="AluSx"
     repeat_region   21654. .21707
                     /rpt_family="AT_rich"
     repeat_region   complement(22726. .22853)
                     /rpt_family="MER5A"
     repeat_region   23951. .24161
                     /rpt_family="MER20"
     repeat_region   complement(24579. .24655)
                     /rpt_family="MIR"
     repeat_region   25020. .25165

Query Match              5.8%;  Score 57.8;  DB 9;  Length 99268;
Best Local Similarity   67.9%;  Pred. No. 0.0032;
```

```
  Matches  112;  Conservative    0;  Mismatches   47;  Indels    6;  Gaps    2;

Qy        829 AGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC--CAAGCAC 886
              ||||||||| |  ||||| |||||  ||||||  |||||     |    | |  |
Db      65407 AGGTCAGAGCATACAAAGTAGCAGGTATGTAGGATGAACAAGGCTAGAATCTACTGTATA 65466

Qy        887 GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAG 946
              |||||||| |||  || |   || |||    |||| ||||||||||| |||||||
Db      65467 ACATGAGGACTGTAGGTAATAAAATTGTACT----GTATGTGGAGTTCATGCTAAATGAG 65522

Qy        947 TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              |   |||||||||||||||||||| || ||||||   ||   ||
Db      65523 TAGATTTTAGCTGCTCTTGCCACAAGAACAAAAAATCTAAGGGAA 65567


RESULT 38
AL136374
LOCUS       AL136374               119853 bp    DNA     linear   PRI 18-JUL-2000
DEFINITION  Human DNA sequence from clone RP1-244G5 on chromosome 1q24.3-25.3,
            complete sequence.
ACCESSION   AL136374
VERSION     AL136374.4  GI:8919204
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119853)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 5, 2000 this sequence version replaced gi:8176662.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP1-244G5 is from the library RPCI-1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP1-244G5.
FEATURES             Location/Qualifiers
     source          1. .119853
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24.3-25.3"
                     /clone="RP1-244G5"
                     /clone_lib="RPCI-1"
     repeat_region   1. .237
                     /note="AluSx repeat: matches 1. .237 of consensus"
     repeat_region   240. .925
                     /note="L2 repeat: matches 1995. .2706 of consensus"
     misc_feature    complement(1507. .1926)
                     /note="match: GSS: Em:AQ146487"
     misc_feature    1822. .2311
                     /note="match: GSS: Em:AQ624491"
     repeat_region   1830. .1956
                     /note="L2 repeat: matches 2616. .2746 of consensus"
     repeat_region   2127. .2181
                     /note="L2 repeat: matches 2656. .2710 of consensus"
     repeat_region   2380. .2691
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   2716. .3083
                     /note="MLT1B repeat: matches 1. .390 of consensus"
     misc_feature    complement(2776. .3233)
                     /note="match: STS: Em:G37725
                     match: GSS: Em:B30422"
     misc_feature    3252. .3690
                     /note="match: GSS: Em:AQ773761"
     repeat_region   4104. .4288
                     /note="MIR repeat: matches 8. .219 of consensus"
     repeat_region   4533. .4756
                     /note="L2 repeat: matches 1992. .2228 of consensus"
     repeat_region   4915. .5144
                     /note="MIR repeat: matches 2. .252 of consensus"
     repeat_region   5284. .5483
                     /note="L2 repeat: matches 2541. .2749 of consensus"
     repeat_region   7056. .7103
                     /note="12 copies 4 mer tcca 85% conserved"
     repeat_region   7216. .7391
                     /note="MIR repeat: matches 73. .262 of consensus"
     repeat_region   8076. .8276
                     /note="MIR repeat: matches 61. .262 of consensus"
     repeat_region   8458. .8494
                     /note="MIR repeat: matches 31. .67 of consensus"
     repeat_region   13489. .13811
                     /note="L2 repeat: matches 1301. .1648 of consensus"
     repeat_region   13837. .13892
                     /note="14 copies 4 mer gtgt 75% conserved"
     repeat_region   14392. .15214
                     /note="MER45B repeat: matches 1. .835 of consensus"
     repeat_region   16024. .16224
                     /note="MIR repeat: matches 34. .248 of consensus"
     repeat_region   16428. .16737
                     /note="AluY repeat: matches 1. .299 of consensus"
     misc_feature    complement(17098. .17761)
                     /note="match: GSS: Em:AQ477207"
     repeat_region   17714. .18068
                     /note="L2 repeat: matches 2155. .2537 of consensus"
     repeat_region   18073. .18710
                     /note="L2 repeat: matches 1049. .1745 of consensus"
     misc_feature    complement(18237. .18836)
                     /note="match: GSS: Em:AQ527362"
     misc_feature    complement(18332. .19056)
                     /note="match: GSS: Em:AQ309729"
     misc_feature    complement(18876. .19056)
                     /note="match: STS: Em:? Em:HSPE58B10"
     misc_feature    18880. .19060
                     /note="match: STS: Em:? Em:HSPE23B05"
     repeat_region   20092. .20163
                     /note="L2 repeat: matches 2673. .2746 of consensus"
     repeat_region   20198. .20725
                     /note="LTR8 repeat: matches 189. .691 of consensus"
     repeat_region   20726. .20992
                     /note="AluSc repeat: matches 1. .299 of consensus"
     repeat_region   20993. .21176
                     /note="LTR8 repeat: matches 1. .189 of consensus"
     repeat_region   21234. .21886
                     /note="LTR1 repeat: matches 1. .660 of consensus"
     repeat_region   22078. .22430
                     /note="LTR16C repeat: matches 3. .385 of consensus"
     repeat_region   22651. .23063
                     /note="MLT1C repeat: matches 50. .466 of consensus"
     repeat_region   23449. .23783
                     /note="MER2 repeat: matches 1. .345 of consensus"
```

```
  misc_feature    complement(24518. .25322)
                  /note="match: GSS: Em:AQ745887"
  repeat_region   24785. .24962
                  /note="MIR repeat: matches 49. .262 of consensus"
  repeat_region   30110. .30149
                  /note="20 copies 2 mer gg 77% conserved"
  repeat_region   30419. .30718
                  /note="AluY repeat: matches 1. .299 of consensus"
  misc_feature    30754. .31346
                  /note="match: GSS: Em:AQ392308"
  repeat_region   30834. .31024
                  /note="MER58A repeat: matches 26. .207 of consensus"
  repeat_region   31661. .31780
                  /note="MIR repeat: matches 2. .115 of consensus"
  repeat_region   31784. .32115
                  /note="L1MA6 repeat: matches 5960. .6294 of consensus"
  repeat_region   32517. .32610
                  /note="MIR repeat: matches 100. .194 of consensus"
  repeat_region   32626. .32939
                  /note="AluSp repeat: matches 1. .313 of consensus"
  repeat_region   33275. .33494
                  /note="L1PA5 repeat: matches 5922. .6145 of consensus"
  repeat_region   33620. .33844
                  /note="MIR repeat: matches 20. .261 of consensus"
  repeat_region   33976. .34701
                  /note="LTR8 repeat: matches 1. .691 of consensus"
  repeat_region   35234. .35696
                  /note="MLT1C repeat: matches 1. .466 of consensus"
  repeat_region   35981. .36020
                  /note="10 copies 4 mer gaga 92% conserved"
  repeat_region   36393. .36458
                  /note="L2 repeat: matches 2682. .2746 of consensus"
  repeat_region   36546. .36711
                  /note="MER5B repeat: matches 2. .173 of consensus"
  misc_feature    complement(37350. .37907)
                  /note="match: GSS: Em:AQ801709"
  repeat_region   37578. .37810
                  /note="AluSg/x repeat: matches 77. .299 of consensus"
  repeat_region   37844. .37947
                  /note="L2 repeat: matches 2589. .2710 of consensus"
  misc_feature    37938. .38503
                  /note="match: GSS: Em:AQ727965"
  misc_feature    complement(38576. .39060)
                  /note="match: GSS: Em:AQ721135"
  repeat_region   39270. .39399
                  /note="MIR repeat: matches 90. .220 of consensus"
  repeat_region   40557. .41199
                  /note="L2 repeat: matches 1882. .2529 of consensus"
  misc_feature    41533. .42009
                  /note="match: GSS: Em:AQ403526"
  repeat_region   42947. .43072
                  /note="MER5A repeat: matches 64. .189 of consensus"
  repeat_region   43714. .43818
                  /note="L2 repeat: matches 2374. .2478 of consensus"
  repeat_region   44053. .44085
                  /note="L2 repeat: matches 2673. .2705 of consensus"
  repeat_region   44219. .44274
                  /note="28 copies 2 mer tt 78% conserved"
  repeat_region   44275. .44569
                  /note="L1PA5 repeat: matches 5849. .6143 of consensus"
  repeat_region   45083. .45140
                  /note="AluSp/q repeat: matches 181. .243 of consensus"
  repeat_region   45141. .45400
                  /note="AluSc repeat: matches 1. .265 of consensus"
  repeat_region   45402. .45515
                  /note="MER5A repeat: matches 51. .187 of consensus"
  repeat_region   46114. .46312
                  /note="L1ME2 repeat: matches 5962. .6154 of consensus"
  repeat_region   46321. .47484
                  /note="L1ME repeat: matches 4307. .5529 of consensus"
  repeat_region   47489. .48076

Query Match          5.8%;  Score 57.8;  DB 9;  Length 119853;
```

```
  Best Local Similarity   56.6%;  Pred. No. 0.0032;
  Matches  107;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        800 GTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTA 859
              ||| ||||     |||| | | |      ||||||||||| |||||| |||| ||||
Db     103264 GTTGGGGGTAGAGAGGAAATGCAGAAATTAGGTCAGAGGATGCAAAGTAGCATATATGTG 103323

Qy        860 GAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGT 919
              | |||||    |       | | |    | | | ||       |
Db     103324 GGATGAACAAGTCTAGAGATCTAATAAACAACATAAGGGCTATAGTTACTAAAAATTGTA 103383

Qy        920 TCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
                .|||||| |  ||  ||| |||| |||   ||| ||||| ||||||||||||||| ||||
Db     103384 CTGTATTTAGGATTATTGCTAAATAAGTAGATTTCAGCTGTTCTTGCCACAAAAACAAAA 103443

Qy        980 AAAAAAAAA 988
                | |||||
Db     103444 CCAGAAAAA 103452


RESULT 39
AL354713
LOCUS       AL354713               138370 bp    DNA     linear   PRI 24-AUG-2002
DEFINITION  Human DNA sequence from clone RP11-498A13 on chromosome 1, complete
            sequence.
ACCESSION   AL354713
VERSION     AL354713.7  GI:22531391
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 138370)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2002 this sequence version replaced gi:14270418.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-498A13 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------.
FEATURES             Location/Qualifiers
     source          1. .138370
```

```
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-498A13"
                     /clone_lib="RPCI-11.2"
BASE COUNT     43651 a  26314 c  27269 g  41136 t
ORIGIN

  Query Match             5.8%;  Score 57.8;  DB 9;  Length 138370;
  Best Local Similarity   56.6%;  Pred. No. 0.0032;
  Matches  107;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        800 GTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTA 859
              ||| ||||     |||| | | |     |||||||||| |||||| ||||   ||||
Db      67409 GTTGGGGGTAGAGAGGAAATGCAGAAATTAGGTCAGAGGATGCAAAGTAGCATATATGTG 67468

Qy        860 GAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGT 919
              | ||||||    |       | |  |    | |  |  ||      |
Db      67469 GGATGAACAAGTCTAGAGATCTAATAAACAACATAAGGGCTATAGTTACTAAAAATTGTA 67528

Qy        920 TCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
              ||||||  |  ||  ||| |||| |||   ||| ||||| ||||||||||||||| ||||
Db      67529 CTGTATTTAGGATTATTGCTAAATAAGTAGATTTCAGCTGTTCTTGCCACAAAAACAAAA 67588

Qy        980 AAAAAAAAA 988
              | |||||
Db      67589 CCAGAAAAA 67597


RESULT 40
AC025375
LOCUS       AC025375               156857 bp    DNA     linear   HTG 25-MAY-2000
DEFINITION  Homo sapiens chromosome 1 clone RP11-26I9 map 1, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC025375
VERSION     AC025375.3  GI:8077072
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 1, clone RP11-26I9
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156857)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7328815.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7958
                Center clone name: 26_I_9
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 149821 bases at least Q40
                Consensus quality: 152963 bases at least Q30
                Consensus quality: 154329 bases at least Q20
                Insert size: 154000; agarose-fp
                Insert size: 155457; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      778: contig of 778 bp in length
            *      779      878: gap of 100 bp
            *      879     1537: contig of 659 bp in length
            *     1538     1637: gap of 100 bp
            *     1638     4250: contig of 2613 bp in length
            *     4251     4350: gap of 100 bp
            *     4351     7895: contig of 3545 bp in length
            *     7896     7995: gap of 100 bp
            *     7996    11871: contig of 3876 bp in length
            *    11872    11971: gap of 100 bp
            *    11972    16256: contig of 4285 bp in length
            *    16257    16356: gap of 100 bp
            *    16357    21822: contig of 5466 bp in length
            *    21823    21922: gap of 100 bp
            *    21923    28321: contig of 6399 bp in length
            *    28322    28421: gap of 100 bp
            *    28422    36684: contig of 8263 bp in length
            *    36685    36784: gap of 100 bp
            *    36785    47214: contig of 10430 bp in length
            *    47215    47314: gap of 100 bp
            *    47315    59012: contig of 11698 bp in length
            *    59013    59112: gap of 100 bp
            *    59113    70244: contig of 11132 bp in length
            *    70245    70344: gap of 100 bp
            *    70345    86759: contig of 16415 bp in length
            *    86760    86859: gap of 100 bp
            *    86860   111698: contig of 24839 bp in length
            *   111699   111798: gap of 100 bp
            *   111799   156857: contig of 45059 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156857
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="1"
                     /clone="RP11-26I9"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .778
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    879. .1537
```

```
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   1638..4250
                    /note="assembly_fragment"
     misc_feature   4351..7895
                    /note="assembly_fragment"
     misc_feature   7996..11871
                    /note="assembly_fragment"
     misc_feature   11972..16256
                    /note="assembly_fragment"
     misc_feature   16357..21822
                    /note="assembly_fragment"
     misc_feature   21923..28321
                    /note="assembly_fragment"
     misc_feature   28422..36684
                    /note="assembly_fragment"
     misc_feature   36785..47214
                    /note="assembly_fragment"
     misc_feature   47315..59012
                    /note="assembly_fragment"
     misc_feature   59113..70244
                    /note="assembly_fragment"
     misc_feature   70345..86759
                    /note="assembly_fragment"
     misc_feature   86860..111698
                    /note="assembly_fragment"
     misc_feature   111799..156857
                    /note="assembly_fragment"
BASE COUNT    47296 a  31366 c  29768 g  47025 t   1402 others
ORIGIN

  Query Match             5.8%;  Score 57.8;  DB 2;  Length 156857;
  Best Local Similarity   56.6%;  Pred. No. 0.0032;
  Matches  107;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        800 GTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTA 859
              ||| ||||      |||| | | |      |||||||||| |||||| ||||   ||||
Db      52955 GTTGGGGGTAGAGAGGAAATGCAGAAATTAGGTCAGAGGATGCAAAGTAGCATATATGTG 53014

Qy        860 GAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGT 919
               | ||||||    |          | |  |    | |  |  ||      |
Db      53015 GGATGAACAAGTCTAGAGATCTAATAAACAACATAAGGGCTATAGTTACTAAAAATTGTA 53074

Qy        920 TCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
                |||||| |  ||  ||| |||| |||    ||| ||||| ||||||||||||| ||||
Db      53075 CTGTATTTAGGATTATTGCTAAATAAGTAGATTTCAGCTGTTCTTGCCACAAAAACAAAA 53134

Qy        980 AAAAAAAAA 988
                | |||||
Db      53135 CCAGAAAAA 53143


RESULT 41
AC079289/c
LOCUS       AC079289               161940 bp    DNA     linear   HTG 20-AUG-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-600J16 map 17, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC079289
VERSION     AC079289.2  GI:22325338
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161940)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-600J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161940)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161940)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 20, 2002 this sequence version replaced gi:9929721.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10440
                Center clone name: 600_J_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 27% of reads
                Sequencing vector: Plasmid; n/a; 73% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 159446 bases at least Q40
                Consensus quality: 160437 bases at least Q30
                Consensus quality: 160863 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 161340; sum-of-contigs
                Quality coverage: 11.9 in Q20 bases; agarose-fp
                Quality coverage: 11.6 in Q2.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *       1    58877: contig of 58877 bp in length
            *   58878    58977: gap of 100 bp
            *   58978    60040: contig of 1063 bp in length
            *   60041    60140: gap of 100 bp
            *   60141    61238: contig of 1098 bp in length
            *   61239    61338: gap of 100 bp
            *   61339    65067: contig of 3729 bp in length
            *   65068    65167: gap of 100 bp
            *   65168   107152: contig of 41985 bp in length
            *  107153   107252: gap of 100 bp
            *  107253   154782: contig of 47530 bp in length
            *  154783   154882: gap of 100 bp
            *  154883   161940: contig of 7058 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161940
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-600J16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .58877
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    58978. .60040
                     /note="assembly_fragment"
     misc_feature    60141. .61238
                     /note="assembly_fragment"
     misc_feature    61339. .65067
                     /note="assembly_fragment"
     misc_feature    65168. .107152
                     /note="assembly_fragment"
     misc_feature    107253. .154782
                     /note="assembly_fragment"
     misc_feature    154883. .161940
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43128 a  36631 c  35390 g  46184 t    607 others
ORIGIN

  Query Match             5.8%;  Score 57.8;  DB 2;  Length 161940;
  Best Local Similarity   67.9%;  Pred. No. 0.0032;
  Matches  112;  Conservative    0;  Mismatches   47;  Indels    6;  Gaps    2;

Qy        829 AGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC--CAAGCAC 886
              ||||||||| |  ||||| |||||  |||||| ||||||    |    | |  |
Db        914 AGGTCAGAGCATACAAAGTAGCAGGTATGTAGGATGAACAAGGCTAGAATCTACTGTATA 855

Qy        887 GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAG 946
              |||||||| |||  || |   || |||     |||| ||||||||||| |||||||
Db        854 ACATGAGGACTGTAGGTAATAAAATTGTACT----GTATGTGGAGTTCATGCTAAATGAG 799

Qy        947 TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              |   |||||||||||||||||||| || ||||||   ||   ||
Db        798 TAGATTTTAGCTGCTCTTGCCACAAGAACAAAAAATCTAAGGGAA 754


RESULT 42
AC107054/c
LOCUS       AC107054               50630 bp    DNA     linear   PRI 29-MAY-2002
DEFINITION  Homo sapiens BAC clone RP11-400H10 from 4, complete sequence.
ACCESSION   AC107054
VERSION     AC107054.5  GI:20429604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50630)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 50630)
  AUTHORS   Wang,C., Kozlowicz,A., Spalding,L. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-400H10
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 50630)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 50630)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 50630)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 50630)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 3, 2002 this sequence version replaced gi:19924180.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0400H10
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-35D5, 2000 bp overlap; the
```

```
          clone sequenced to the right is RP11-642E20, 2000 bp overlap.
          Actual start of this clone is at base position 97704 of RP11-35D5.
FEATURES             Location/Qualifiers
     source          1. .50630
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-400H10"
                     /clone_lib="RPCI-11"
     repeat_region   151. .630
                     /rpt_family="MaLR"
     repeat_region   1007. .1305
                     /rpt_family="Alu"
     repeat_region   2780. .3061
                     /rpt_family="L2"
     repeat_region   3241. .3545
                     /rpt_family="L2"
     repeat_region   4683. .6142
                     /rpt_family="L1"
     repeat_region   6143. .6446
                     /rpt_family="Alu"
     repeat_region   6447. .7296
                     /rpt_family="L1"
     repeat_region   7297. .7669
                     /rpt_family="MaLR"
     repeat_region   7670. .8151
                     /rpt_family="L1"
     repeat_region   8411. .8521
                     /rpt_family="MIR"
     repeat_region   8732. .8767
                     /rpt_family="AT_rich"
     repeat_region   10181. .10668
                     /rpt_family="L2"
     repeat_region   10669. .10936
                     /rpt_family="Alu"
     repeat_region   10937. .11264
                     /rpt_family="L2"
     repeat_region   12145. .12226
                     /rpt_family="L2"
     repeat_region   13486. .13686
                     /rpt_family="MIR"
     repeat_region   14333. .14356
                     /rpt_family="(CAAAA)n"
     repeat_region   14395. .14548
                     /rpt_family="MIR"
     repeat_region   14678. .14849
                     /rpt_family="MIR"
     repeat_region   15282. .15369
                     /rpt_family="L2"
     repeat_region   15487. .15510
                     /rpt_family="AT_rich"
     repeat_region   15511. .15572
                     /rpt_family="Alu"
     repeat_region   15607. .15753
                     /rpt_family="L2"
     repeat_region   15982. .16156
                     /rpt_family="L1"
     repeat_region   16164. .16390
                     /rpt_family="MaLR"
     repeat_region   16406. .16465
                     /rpt_family="MER2_type"
     repeat_region   16466. .18051
                     /rpt_family="L1"
     repeat_region   18052. .18369
                     /rpt_family="Alu"
     repeat_region   18370. .18939
                     /rpt_family="L1"
     repeat_region   18960. .19089
                     /rpt_family="L1"
     repeat_region   19231. .19323
                     /rpt_family="L1"
     repeat_region   19322. .20169
                     /rpt_family="L1"
     repeat_region   20170. .20479
                     /rpt_family="Alu"
     repeat_region   20480. .20630
                     /rpt_family="L1"
     repeat_region   20800. .21207
                     /rpt_family="MaLR"
     repeat_region   21687. .21731
                     /rpt_family="AT_rich"
     repeat_region   21782. .21840
                     /rpt_family="L2"
     repeat_region   23361. .23671
                     /rpt_family="Alu"
     repeat_region   24503. .24525
                     /rpt_family="AT_rich"
     repeat_region   26923. .27223
                     /rpt_family="Alu"
     repeat_region   27248. .27504
                     /rpt_family="Alu"
     repeat_region   27930. .27951
                     /rpt_family="(T)n"
     repeat_region   28083. .28152
                     /rpt_family="Mariner"
     repeat_region   28172. .28215
                     /rpt_family="AT_rich"
     repeat_region   29024. .29147
                     /rpt_family="L2"
     repeat_region   29148. .29264
                     /rpt_family="(TA)n"
     repeat_region   29620. .29682
                     /rpt_family="L2"
     repeat_region   30174. .30291
                     /rpt_family="L2"
     repeat_region   30570. .30623
                     /rpt_family="CR1"
     repeat_region   30632. .30692
                     /rpt_family="MER1_type"
     repeat_region   30776. .30821
                     /rpt_family="AT_rich"
     repeat_region   31291. .31420
                     /rpt_family="L1"
     repeat_region   33262. .33396
                     /rpt_family="MIR"

  Query Match             5.8%;  Score 57.4;  DB 9;  Length 50630;
  Best Local Similarity   63.1%;  Pred. No. 0.0041;
  Matches  123;  Conservative    0;  Mismatches   66;  Indels    6;  Gaps    2;

Qy        793 AAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAG 852
              ||  |  |||   ||||      |  ||| |    |  |||||||  |||||  |||||
Db       4996 AATGGTGGTTATAGGGGCAATTAAATGGGGAGATGTCAGCCAGAGGATACAAAGTAGCAG 4937

Qy        853 CCATGTAGAATGAACCGTCCAGAGAGCCAA--GCACGGCAGAGGACTGCAGGCCATCAGC 910
                ||||||  |||  |  |||  ||||  | ||          |||||||  |||  || |
Db       4936 ATATGTAGGATGGAAAGTCTGGAGATCTAATGTACAACATGAGGACTATAGGTAATAAAA 4877

Qy        911 GTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACA 970
               ||  |    |  |||  |  |  |||||||  ||||||||   ||||||||||| ||||||||
Db       4876 TTGTA----TTCTATGTAGGATTCATGCTAAATGAGTAGATTTTAGCTGCTGTTGCCACA 4821

Qy        971 AAAAAAAAAAAAAAA 985
              ||||  |||  ||||||
Db       4820 AAAACAAACAAAAAA 4806


RESULT 43
AL357129
LOCUS       AL357129              145380 bp    DNA     linear   PRI 21-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-114A21 on chromosome
            Xq21.31-22.1 Contains ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL357129
VERSION     AL357129.11  GI:10186549
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145380)
  AUTHORS   Chapman,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2000 this sequence version replaced gi:9909073.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP11-114A21 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-114A21 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-114A21 is at 1 in this sequence.
            The true left end of clone RP1-122K23 is at 145281 in this
            sequence. The true left end of clone RP11-554A24 is at 80790 in
            this sequence. The true right end of clone RP11-361B11 is at 54224
            in this sequence.
FEATURES             Location/Qualifiers
     source          1. .145380
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q21.31-22.1"
                     /clone="RP11-114A21"
                     /clone_lib="RPCI-11.1"
     repeat_region   1. .71
                     /note="L1P3 repeat: matches 10. .80 of consensus"
     misc_feature    20. .685
                     /note="match: GSS: Em:AQ343933"
     repeat_region   31. .79
                     /note="L1P repeat: matches 4. .52 of consensus"
     repeat_region   450. .492
                     /note="MER63 repeat: matches 2. .44 of consensus"
     repeat_region   847. .993
                     /note="MER20 repeat: matches 29. .174 of consensus"
     repeat_region   1591. .1614
                     /note="12 copies 2 mer tc 95% conserved"
     repeat_region   3441. .3588
                     /note="2 copies 74 mer 92% conserved"
     repeat_region   3742. .3801
                     /note="L2 repeat: matches 2643. .2701 of consensus"
     repeat_region   4195. .4440
                     /note="MLT1A1 repeat: matches 118. .365 of consensus"
     repeat_region   4442. .4495
                     /note="HERV17 repeat: matches 1. .54 of consensus"
     repeat_region   4496. .5265
                     /note="LTR17 repeat: matches 1. .780 of consensus"
     repeat_region   5271. .6471
                     /note="L1PBa repeat: matches 1037. .2339 of consensus"
     repeat_region   6464. .6757
                     /note="L1PBa repeat: matches -712. .-411 of consensus"
     repeat_region   6786. .7118
                     /note="L1PBa repeat: matches -176. .1104 of consensus"
     repeat_region   7770. .8014
                     /note="MLT1H repeat: matches 38. .289 of consensus"
     repeat_region   8097. .8234
                     /note="MLT1H repeat: matches 410. .547 of consensus"
     repeat_region   8310. .8468
                     /note="L2 repeat: matches 2563. .2723 of consensus"
     repeat_region   8484. .8639
                     /note="L2 repeat: matches 1060. .1224 of consensus"
     repeat_region   9501. .10717
                     /note="L1P3 repeat: matches 4. .924 of consensus"
     repeat_region   10749. .11059
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   11101. .13160
                     /note="L1P repeat: matches 948. .2998 of consensus"
     repeat_region   13159. .15022
                     /note="L1PA8 repeat: matches 4290. .6162 of consensus"
     repeat_region   15671. .15793
                     /note="L2 repeat: matches 2577. .2710 of consensus"
     repeat_region   16891. .17018
                     /note="MIR repeat: matches 14. .144 of consensus"
     repeat_region   18274. .18889
                     /note="L2 repeat: matches 1656. .2319 of consensus"
     repeat_region   18868. .18953
                     /note="MIR repeat: matches 169. .254 of consensus"
     repeat_region   19189. .19961
                     /note="L1MA2 repeat: matches 5536. .6308 of consensus"
     repeat_region   19962. .21580
                     /note="L1PA13 repeat: matches 4433. .6073 of consensus"
     repeat_region   21585. .24399
                     /note="L1PA13 repeat: matches 1532. .4323 of consensus"
     repeat_region   24420. .25728
                     /note="L1M1 repeat: matches -301. .1172 of consensus"
     repeat_region   25721. .25834
                     /note="L1M1 repeat: matches -1389. .-1269 of consensus"
     repeat_region   26112. .26285
                     /note="MER58A repeat: matches 45. .224 of consensus"
     repeat_region   26387. .26742
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   26857. .27201
                     /note="L1ME repeat: matches 5514. .5867 of consensus"
     repeat_region   27851. .27978
                     /note="MER47 repeat: matches 2215. .2323 of consensus"
     repeat_region   27979. .28352
                     /note="THE1C repeat: matches 5. .371 of consensus"
     repeat_region   28353. .28503
                     /note="MER47 repeat: matches 2058. .2215 of consensus"
     repeat_region   28529. .28645
                     /note="MIR repeat: matches 119. .245 of consensus"
     repeat_region   28672. .29928
                     /note="MER57-internal repeat: matches 1731. .2937 of
                     consensus"
     repeat_region   29688. .30062
                     /note="HUERS-P3b repeat: matches 2227. .2560 of consensus"
     repeat_region   30076. .30409
                     /note="MER57-internal repeat: matches 1306. .1630 of
                     consensus"
     misc_feature    30650. .30978
                     /note="match: GSS: Em:B38708"
     misc_feature    30654. .30871
                     /note="match: GSS: Em:AQ589046"
     repeat_region   31264. .31806
                     /note="LTR8 repeat: matches 159. .691 of consensus"
     repeat_region   31816. .32111
                     /note="AluSx repeat: matches 4. .299 of consensus"
     repeat_region   32145. .32281
```

```
                     /note="PABL_B repeat: matches 204. .351 of consensus"
     repeat_region   32282. .32586
                     /note="AluSp repeat: matches 7. .310 of consensus"
     repeat_region   32587. .32734
                     /note="PABL_B repeat: matches 102. .204 of consensus"
     repeat_region   32735. .33047
                     /note="AluSp repeat: matches 3. .313 of consensus"
     repeat_region   33048. .33142
                     /note="PABL_B repeat: matches 1. .102 of consensus"
     repeat_region   33207. .33283
                     /note="HUERS-P3b repeat: matches 7308. .7378 of consensus"
     misc_feature    complement(33673. .34083)
                     /note="match: GSS: Em:AQ034800"
     misc_feature    complement(34279. .34523)
                     /note="match: STS: Em:G58090
                     match: GSS: Em:AQ384354"
     misc_feature    34950. .35388
                     /note="match: GSS: Em:AQ664261"
     repeat_region   35211. .35546
                     /note="HERVHC2 repeat: matches 4748. .5089 of consensus"
     repeat_region   36077. .36508
                     /note="HUERS-P3 repeat: matches 4694. .5118 of consensus"
     repeat_region   36598. .37624
                     /note="HERVE repeat: matches 2672. .3008 of consensus"
     repeat_region   38168. .38291
                     /note="HERV9 repeat: matches 2732. .2840 of consensus"
     repeat_region   38939. .38974
                     /note="U2 repeat: matches 1. .36 of consensus"
     repeat_region   40296. .40966
                     /note="PABL_B repeat: matches 1. .667 of consensus"
     repeat_region   40976. .41346
                     /note="THE1C repeat: matches 15. .369 of consensus"
     repeat_region   41348. .41495
                     /note="MER47 repeat: matches 2062. .2210 of consensus"
     repeat_region   41525. .41641
                     /note="MIR repeat: matches 119. .245 of consensus"
     repeat_region   42737. .42875
                     /note="MIR repeat: matches 110. .248 of consensus"
     misc_feature    43217. .43703
                     /note="match: GSS: Em:AQ522941"
     repeat_region   43929. .44074
                     /note="L2 repeat: matches 2595. .2748 of consensus"
     repeat_region   44087. .44550
                     /note="L1MC/D repeat: matches 5354. .5823 of consensus"
     repeat_region   44551. .44698
                     /note="L1ME1 repeat: matches 5685. .5825 of consensus"
     repeat_region   44730. .44791
                     /note="L1MA4 repeat: matches 6234. .6295 of consensus"
     repeat_region   44794. .45004

  Query Match             5.8%;  Score 57.4;  DB 9;  Length 145380;
  Best Local Similarity   55.9%;  Pred. No. 0.004;
  Matches  109;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        795 GAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCC 854
              | || | | | || |       | | | | || |||||||||   |||||| |||||
Db      86358 GTGGGGATGGAGGTGGGACTAAATGAGGAGAGGTAGGTCAGAGCTTGCAAAGTAGCAGAT 86417

Qy        855 ATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGC 914
              | |||| ||||||     |      | |       | | ||    |       |
Db      86418 AAGTAGGATGAACAAGTCTAGAGATCTAATGTACAATATGAGAACTATAGCTTTAAAAAA 86477

Qy        915 ACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
              | |||      ||||  ||| ||  ||||||||    |||||||||||||||| ||||||
Db      86478 ATTGTATTGTATTGGGTTTCCTGTTAAATGAGTAGCTTTTAGCTGCTCTTGCTACAAAAT 86537

Qy        975 AAAAAAAAAAAAAAA 989
              || || ||| ||||
Db      86538 CAACAACAAAGAAAA 86552


RESULT 44
AC021021
```

```
LOCUS       AC021021              169202 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-557N1 from 2, complete sequence.
ACCESSION   AC021021
VERSION     AC021021.9  GI:14140347
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169202)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 169202)
  AUTHORS   VanBrunt,A., Hawkins,M., Haakenson,W. and Mullen,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-557N1
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 169202)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 169202)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 169202)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 17, 2001 this sequence version replaced gi:13399429.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0557N01
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
```

(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC021789; the clone sequenced to the right is AC021688. Actual start of this clone is at base position 1 of RP11-557N1; actual end is at base position 169202 of RP11-557N1.

The sequence H_NH0557N01 contains a dinucleotide (GA) repeat from base position 29551 to 29868 where the sequence fidelity cannot be guaranteed. Assembly of the database is consistent with digest information.

FEATURES Location/Qualifiers

source 1. .169202
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-557N1"
/clone_lib="RPCI-11"

repeat_region 196. .300
/rpt_family="Mariner"

repeat_region 407. .730
/rpt_family="MER121"

repeat_region 891. .934
/rpt_family="AT_rich"

repeat_region 1399. .1428
/rpt_family="AT_rich"

repeat_region 1550. .1577
/rpt_family="AT_rich"

repeat_region 1739. .1895
/rpt_family="MIR"

repeat_region 3511. .3635
/rpt_family="L2"

repeat_region 3700. .4079
/rpt_family="ERV1"

repeat_region 4106. .4201
/rpt_family="L2"

repeat_region 4947. .5133
/rpt_family="MIR"

repeat_region 5960. .5996
/rpt_family="polypyrimidine"

repeat_region 6720. .6772
/rpt_family="MIR"

repeat_region 6809. .6889
/rpt_family="MIR"

repeat_region 7122. .7348
/rpt_family="L1"

repeat_region 7349. .7392
/rpt_family="(GTTTG)n"

repeat_region 7393. .7802
/rpt_family="L1"

repeat_region 7803. .7856
/rpt_family="(TA)n"

repeat_region 7867. .7984
/rpt_family="L1"

repeat_region 8738. .8860
/rpt_family="MIR"

repeat_region 8863. .9145
/rpt_family="Alu"

repeat_region 9426. .9714
/rpt_family="Alu"

repeat_region 9715. .9758
/rpt_family="AT_rich"

repeat_region 9903. .9977
/rpt_family="L2"

repeat_region 10028. .10056
/rpt_family="(TG)n"

repeat_region 10404. .10689
/rpt_family="Alu"

repeat_region 12107. .12278
/rpt_family="MaLR"

repeat_region 13854. .13946
/rpt_family="AcHobo"

repeat_region 14020. .14183
/rpt_family="MIR"

repeat_region 14531. .14841
/rpt_family="Alu"

repeat_region 14860. .15243
/rpt_family="MER2_type"

repeat_region 16032. .16338
/rpt_family="Alu"

repeat_region 17440. .17564
/rpt_family="CR1"

repeat_region 19909. .19941
/rpt_family="AT_rich"

repeat_region 20077. .20139
/rpt_family="MIR"

repeat_region 20386. .20982
/rpt_family="ERV1"

repeat_region 21156. .21524
/rpt_family="MaLR"

repeat_region 22038. .22097
/rpt_family="MIR"

repeat_region 22230. .22331
/rpt_family="MIR"

repeat_region 24135. .24157
/rpt_family="AT_rich"

repeat_region 25901. .25968
/rpt_family="MIR"

repeat_region 27874. .28003
/rpt_family="MIR"

repeat_region 28292. .28426
/rpt_family="MIR"

repeat_region 28611. .28718
/rpt_family="L2"

repeat_region 28774. .28835
/rpt_family="(TTAAA)n"

repeat_region 29147. .29261
/rpt_family="GA-rich"

repeat_region 29285. .29567
/rpt_family="Alu"

repeat_region 29568. .29642
/rpt_family="polypurine"

repeat_region 29644. .29831
/rpt_family="GA-rich"

repeat_region 29832. .29866
/rpt_family="(GGGA)n"

repeat_region 29991. .30283
/rpt_family="Alu"

repeat_region 30686. .31803
/rpt_family="L1"

repeat_region 31804. .32105
/rpt_family="Alu"

repeat_region 32106. .32750
/rpt_family="L1"

Query Match 5.8%; Score 57.4; DB 9; Length 169202;
Best Local Similarity 64.2%; Pred. No. 0.004;
Matches 120; Conservative 0; Mismatches 61; Indels 6; Gaps 2;

```
Qy      805 GGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATG 864
            | |||   |||| |  |   ||| |||||||| |||   |||| ||||   ||| || |||
Db    79581 GTGGGAGAAGGAAACTGGCGGGGTAGGTCAGGGGATAGAAAGTAGCAAATATGAAGGATG 79640

Qy      865 AACCGTCCAGAGAGCC--AAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCG 922
            |||    | |    ||  ||| |  || |  ||| || |  || | |  |  ||   |
Db    79641 AACAAGGCTGGAGACCTACAGCCCAACATACCACTACAAGTAATAAACTCGAGCT----G 79696

Qy      923 TATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAA 982
            |||||||  |||||||  ||||  |  ||||||||||||||||||||  ||| |||||||
Db    79697 TATTTGGGATTCATGCTGAATGGCTAGATTTTAGCTGCTCTTGCCACACAAAGAAAAAAA 79756

Qy      983 AAAAAAA 989
                ||| ||
```

```
Db      79757 TGAATAA 79763


RESULT 45
AC145171/c
LOCUS       AC145171               175947 bp    DNA     linear   HTG 12-JUN-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-349F16, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC145171
VERSION     AC145171.1  GI:31621312
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175947)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175947)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 540517
            Center clone name: RPCI-11_349F16
            -------
            Summary Statistics
            Consensus quality: 173647 bases at least Q40
            Consensus quality: 174869 bases at least Q30
            Consensus quality: 175338 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 175747; sum-of-contigs estimation
            Quality coverage:  9.21 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.17 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    45369: contig of 45369 bp in length
            *    45370    45469: gap of unknown length
            *    45470    94891: contig of 49422 bp in length
            *    94892    94991: gap of unknown length
            *    94992   175947: contig of 80956 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175947
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-349F16"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    54782 a  31550 c  31265 g  58150 t    200 others
ORIGIN

  Query Match             5.8%;  Score 57.4;  DB 2;  Length 175947;
  Best Local Similarity   55.9%;  Pred. No. 0.004;
  Matches  109;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy        795 GAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCC 854
              | || | | | || |      |  | | |  || |||||||||   |||||| |||||
Db      39798 GTGGGGATGGAGGTGGGACTAAATGAGGAGAGGTAGGTCAGAGCTTGCAAAGTAGCAGAT 39739

Qy        855 ATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGC 914
              | |||| ||||||   |  | |   | | ||  |   |
Db      39738 AAGTAGGATGAACAAGTCTAGAGATCTAATGTACAATATGAGAACTATAGCTTTAAAAAA 39679

Qy        915 ACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
              | |||      ||||  ||| ||  |||||||   ||||||||||||||| ||||||
Db      39678 ATTGTATTGTATTGGGTTTCCTGTTAAATGAGTAGCTTTTAGCTGCTCTTGCTACAAAAT 39619

Qy        975 AAAAAAAAAAAAAAA 989
              || || ||| ||||
Db      39618 CAACAACAAAGAAAA 39604


RESULT 46
AC013408
LOCUS       AC013408               151841 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-469G18 from 2, complete sequence.
ACCESSION   AC013408
VERSION     AC013408.7  GI:15920142
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151841)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 151841)
  AUTHORS   Holmes,A., Hawkins,M. and Kozlowicz,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-469G18
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 151841)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 151841)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 151841)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 4, 2001 this sequence version replaced gi:13992768.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0469G18
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.
```

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-90C9, 2000 bp overlap; the clone sequenced to the right is RP11-22L19, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-469G18; actual end is at base position 14496 of RP11-22L19.

Data from AC079146 was used to finish this clone, AC013408.
Polymorphisms have been identified between AC079146 and AC013408.

FEATURES	Location/Qualifiers
source	1. .151841
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-469G18"
	/clone_lib="RPCI-11"
repeat_region	2. .1209
	/rpt_family="ERVK"
repeat_region	1210. .2222
	/rpt_family="ERVK"
repeat_region	3426. .3844
	/rpt_family="L1"
repeat_region	3559. .3593
	/rpt_family="(CA)n"
repeat_region	5300. .5476
	/rpt_family="MIR"
repeat_region	5483. .5601
	/rpt_family="AcHobo"
repeat_region	6149. .6256
	/rpt_family="L2"
repeat_region	6526. .6700
	/rpt_family="MaLR"
repeat_region	9037. .9184
	/rpt_family="(TA)n"
repeat_region	9834. .9864
	/rpt_family="(GA)n"
repeat_region	12527. .12597
	/rpt_family="(TA)n"
repeat_region	13739. .13760
	/rpt_family="AT_rich"
repeat_region	13964. .13999
	/rpt_family="AT_rich"
repeat_region	14037. .14242
	/rpt_family="MaLR"
repeat_region	14214. .14240
	/rpt_family="(TC)n"
repeat_region	14483. .14787
	/rpt_family="Alu"
repeat_region	15016. .15042
	/rpt_family="(TAGA)n"
repeat_region	15269. .15555
	/rpt_family="Alu"
repeat_region	17128. .17147
	/rpt_family="(TTTTTA)n"
repeat_region	17131. .17430
	/rpt_family="Alu"
repeat_region	17445. .17472
	/rpt_family="(T)n"
repeat_region	17760. .17782
	/rpt_family="AT_rich"
repeat_region	17859. .17931
	/rpt_family="MIR"
repeat_region	18428. .18484
	/rpt_family="MaLR"
repeat_region	18474. .18496
	/rpt_family="AT_rich"
repeat_region	18485. .19089
	/rpt_family="L1"
repeat_region	19090. .19431
	/rpt_family="MaLR"
repeat_region	19465. .19547
	/rpt_family="GA-rich"
repeat_region	22482. .22512
	/rpt_family="AT_rich"
repeat_region	22637. .22734
	/rpt_family="L1"
repeat_region	29040. .29078
	/rpt_family="(CA)n"
repeat_region	29467. .29576
	/rpt_family="MIR"
repeat_region	29686. .29874
	/rpt_family="MIR"
repeat_region	30573. .30990
	/rpt_family="ERVL"
repeat_region	30605. .30627
	/rpt_family="AT_rich"
repeat_region	31817. .31898
	/rpt_family="(TA)n"
repeat_region	31979. .32006
	/rpt_family="AT_rich"
repeat_region	32107. .32423
	/rpt_family="Alu"
repeat_region	32397. .32427
	/rpt_family="(TAAA)n"
repeat_region	33618. .33656
	/rpt_family="AT_rich"
repeat_region	35713. .36026
	/rpt_family="MER1_type"
repeat_region	35814. .35845
	/rpt_family="AT_rich"
repeat_region	35904. .35947
	/rpt_family="AT_rich"
repeat_region	37385. .37417
	/rpt_family="AT_rich"
repeat_region	37594. .38042
	/rpt_family="L1"
repeat_region	37733. .37788
	/rpt_family="(TG)n"
repeat_region	38010. .38313
	/rpt_family="L1"
repeat_region	38406. .38429
	/rpt_family="(TTG)n"
repeat_region	38847. .38946
	/rpt_family="MIR"
repeat_region	38947. .39247
	/rpt_family="Alu"
repeat_region	39248. .39385
	/rpt_family="MIR"
repeat_region	39419. .39650
	/rpt_family="L1"
repeat_region	41061. .41384
	/rpt_family="L1"
repeat_region	41414. .41583
	/rpt_family="MIR"

Query Match 5.7%; Score 56.6; DB 9; Length 151841;
Best Local Similarity 56.0%; Pred. No. 0.0062;
Matches 107; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
Qy        786 CTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAA 845
              ||| ||   ||  | ||| ||     | |   |   || |||||| | ||| ||||
Db     124592 CTGGTTGCTATGGGTATGGGGTGGGGAGGAAATGAAGACAGGTAGGTCAAAAGACACAAA 124651

Qy        846 GCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCA 905
              |||||  |||||||||||||    | ||    | | |    || ||   ||    |
Db     124652 ATAGCAGGTATGTAGAATGAACAAGTCTGAACATCTAATAAACATGATAACCATAGTTAA 124711

Qy        906 TCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTG 965
              | |    ||| | |   ||      ||  ||||| ||| ||||||||| |||||
Db     124712 TTAAATGGCATTATATGAGGTTTTTTAAAATAAAAAATAAGTTGGTTTTAGCTGTTCTTG 124771

Qy        966 CCACAAAAAAA 976
              |||||  |||
Db     124772 TCACAAGTAAA 124782
```

RESULT 47
AC025522
LOCUS AC025522 173270 bp DNA linear PRI 05-FEB-2003
DEFINITION Homo sapiens chromosome 8, clone RP11-599P7, complete sequence.
ACCESSION AC025522
VERSION AC025522.7 GI:28209714
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173270)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-599P7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173270)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 173270)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 173270)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 5, 2003 this sequence version replaced gi:18652546.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L8317
Center clone name: 599_P_7
--------------.

```
FEATURES             Location/Qualifiers
     source          1. .173270
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-599P7"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   238. .425
                     /rpt_family="MIR"
     repeat_region   1545. .1691
                     /rpt_family="MIR"
     repeat_region   2232. .2562
                     /rpt_family="L1PB4"
     unsure          2495. .2649
                     /note="Single clone coverage"
     repeat_region   2614. .2734
                     /rpt_family="MIR"
     repeat_region   3063. .3212
                     /rpt_family="L1MC/D"
     repeat_region   3213. .3539
                     /rpt_family="AluJb"
     repeat_region   3540. .3845
                     /rpt_family="L1MC/D"
```

```
repeat_region   complement(4673. .4745)
                /rpt_family="MIR"
repeat_region   complement(4765. .4938)
                /rpt_family="MIR"
repeat_region   6220. .6268
                /rpt_family="AT_rich"
repeat_region   complement(6368. .6543)
                /rpt_family="MIR"
repeat_region   7369. .7664
                /rpt_family="AluSx"
repeat_region   complement(7959. .8673)
                /rpt_family="LTR1B"
repeat_region   9636. .9686
                /rpt_family="MIR"
repeat_region   10618. .10638
                /rpt_family="AT_rich"
repeat_region   complement(11204. .11513)
                /rpt_family="AluSg"
repeat_region   11624. .12119
                /rpt_family="MLT1D"
repeat_region   complement(12125. .12194)
                /rpt_family="MIR"
repeat_region   12717. .12757
                /rpt_family="AT_rich"
repeat_region   complement(13082. .13389)
                /rpt_family="AluY"
repeat_region   15244. .15324
                /rpt_family="AT_rich"
repeat_region   complement(15341. .15853)
                /rpt_family="L1M2"
repeat_region   15861. .15969
                /rpt_family="FRAM"
repeat_region   15970. .15992
                /rpt_family="(TTG)n"
repeat_region   complement(15997. .16293)
                /rpt_family="AluSp"
repeat_region   complement(16318. .16379)
                /rpt_family="L1M4"
repeat_region   16417. .16534
                /rpt_family="FLAM_C"
repeat_region   16672. .16697
                /rpt_family="AT_rich"
repeat_region   complement(16770. .16958)
                /rpt_family="L1MC4"
repeat_region   complement(16962. .17095)
                /rpt_family="L1M2"
repeat_region   complement(17096. .17477)
                /rpt_family="L1MC4"
repeat_region   17514. .17534
                /rpt_family="AT_rich"
repeat_region   complement(17535. .17879)
                /rpt_family="L1MC4a"
repeat_region   17880. .17910
                /rpt_family="(TA)n"
repeat_region   complement(17911. .18043)
                /rpt_family="L1MC4a"
repeat_region   complement(18667. .18983)
                /rpt_family="L1M2"
repeat_region   19055. .19102
                /rpt_family="(TG)n"
repeat_region   19398. .19494
                /rpt_family="T-rich"
repeat_region   20011. .20043
                /rpt_family="AT_rich"
repeat_region   complement(20337. .20507)
                /rpt_family="HAL1b"
unsure          complement(20385. .20390)
                /note="<30 qual SNGL region"
repeat_region   20966. .20995
                /rpt_family="(TG)n"
repeat_region   20995. .21020
                /rpt_family="(GA)n"
repeat_region   22075. .22119
                /rpt_family="AT_rich"
```

```
  Query Match             5.7%;  Score 56.6;  DB 9;  Length 173270;
  Best Local Similarity   66.9%;  Pred. No. 0.0062;
  Matches  113;  Conservative    0;  Mismatches   49;  Indels    7;  Gaps    2;

Qy        814 GGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC---CGT 870
              |  | ||| |  | ||||||||||| ||||| || |  |||||| ||||||   |
Db      42074 GAAATGGGGAGATGTAGGTCAGAGGATACAAAGTAGAATATATGTAGGATGAACAAGTGA 42133

Qy        871 CCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGA 930
              |||       |       ||||| |  ||| || |  ||||||    |||||||||
Db      42134 AGAGATCTAATATACGATATGAGGATTATAGGTAATAACATTGCACT----GTATTTGGA 42189

Qy        931 GTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
              ||||||| ||||||||   ||||||||| ||||||||||||||||||
Db      42190 ATTCATGCTAAATGAGTAGATTTTAGCTGTTCTTGCCACAAAAAAAAAA 42238


RESULT 48
AC016256
LOCUS       AC016256               164405 bp    DNA     linear   PRI 25-JUL-2001
DEFINITION  Homo sapiens 12 BAC RP11-98E6 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC016256
VERSION     AC016256.19  GI:14670071
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164405)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R. and Gibbs,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 164405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 164405)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 11, 2001 this sequence version replaced gi:12656764.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES                Location/Qualifiers
     source             1. .164405
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="12"
                        /clone="RP11-98E6"
     repeat_region      complement(1. .371)
                        /rpt_family="MLT1F1"
     misc_feature       3. .5166
                        /note="Overlaps bases 186760. .191923 of clone AC087886"
                        /function="Overlaps with adjacent clone AC087886"
     repeat_region      complement(439. .637)
                        /rpt_family="L1MA8"
     repeat_region      2142. .2281
                        /rpt_family="MIR"
     repeat_region      complement(2827. .2961)
                        /rpt_family="MIR"
     repeat_region      3943. .4314
                        /rpt_family="L2"
     repeat_region      5502. .5527
                        /rpt_family="AT_rich"
     repeat_region      6069. .6615
                        /rpt_family="L1ME3A"
     repeat_region      6830. .6911
                        /rpt_family="L2"
     repeat_region      7015. .7068
                        /rpt_family="AT_rich"
     repeat_region      7069. .7153
                        /rpt_family="(TA)n"
     repeat_region      7163. .7191
                        /rpt_family="(TG)n"
     repeat_region      complement(7192. .7324)
                        /rpt_family="L2"
     repeat_region      complement(7387. .7463)
                        /rpt_family="L2"
     repeat_region      7825. .8933
                        /rpt_family="L1PA13"
     repeat_region      complement(8937. .9171)
                        /rpt_family="L2"
     repeat_region      9494. .9629
                        /rpt_family="L1"
     repeat_region      complement(9700. .10075)
                        /rpt_family="L1MA9"
     repeat_region      10071. .10234
                        /rpt_family="L1MA9"
     repeat_region      10294. .10581
                        /rpt_family="L1M4"
     repeat_region      10582. .11009
                        /rpt_family="Tigger2a"
     repeat_region      11010. .11303
                        /rpt_family="L1M4"
     repeat_region      11500. .11670
                        /rpt_family="L1M3"
     repeat_region      11701. .12100
                        /rpt_family="L1M3"
     repeat_region      12185. .13027
                        /rpt_family="L1M3"
     repeat_region      complement(13224. .13505)
                        /rpt_family="L1PA10"
     repeat_region      13929. .13957
                        /rpt_family="AT_rich"
     repeat_region      complement(13958. .15549)
                        /rpt_family="L1PA3"
     misc_feature       14218
                        /note="A or G"
                        /function="Polymorphic site"
     misc_feature       14264
                        /note="C or A"
                        /function="Polymorphic site"
     misc_feature       14283
                        /note="C or A"
                        /function="Polymorphic site"
     misc_feature       14288
                        /note="C or T"
                        /function="Polymorphic site"
     STS                14300. .14431
                        /standard_name="G19948"
                        /db_xref="dbSTS:32826"
     misc_feature       14356
                        /note="C or T"
                        /function="Polymorphic site"
     misc_feature       14362
                        /note="C or G"
                        /function="Polymorphic site"
     misc_feature       14397. .14398
                        /note="CA or TG"
```

```
                     /function="Polymorphic site"
     misc_feature    14426

  Query Match             5.7%;  Score 56.4;  DB 9;  Length 164405;
  Best Local Similarity   58.1%;  Pred. No. 0.007;
  Matches  137;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    2;

Qy        757 CCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGA 816
              ||||| | || ||  |  |  | | | | ||| |   || ||| ||  |
Db     143164 CCTTGAATATATAGAGATAGAGAAGAAAGCAGTGGATACAGGTGGAGAGGTGGTCCAAAA 143223

Qy        817 TAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAG 876
              |||||  | || | |||||  |||| | || |||||| ||||||    |
Db     143224 ATGGGAAGATGTAGTTAAGAGGATACAAATTA--AGTTATGTAGGATGAACAAGTCTAGA 143281

Qy        877 AGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCAT 936
              | | |  |   | | ||  |    || |      |  | || | ||  |||||
Db     143282 GGTCTAAGATACAATATGAGGACTATAAATAATAAAATTGTACT-GTGTCTGTGATTCAT 143340

Qy        937 GCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              || ||||||||  |||||||||||||||||||||||  ||||||||||||||||
Db     143341 GCTAAATGAGTCGATTTTAGCTGCTCTTGCCACAAAACCAAAAAAAAAAAAAAAAA 143396


RESULT 49
AC025258/c
LOCUS       AC025258               166668 bp    DNA     linear   HTG 15-OCT-2001
DEFINITION  Homo sapiens chromosome 12 clone RP11-398O17, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC025258
VERSION     AC025258.15  GI:14547352
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166668)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166668)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 25, 2001 this sequence version replaced gi:12621178.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HALL
                Center clone name: RP11-398O17
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 27% of reads
                Chemistry: Dye-terminator Big Dye: 73% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 177529 bases at least Q40
                Consensus quality: 182522 bases at least Q30
                Consensus quality: 185426 bases at least Q20
                Estimated insert size: 170888; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   149576: contig of 149576 bp in length
            *  149577   149676: gap of unknown length
            *  149677   166668: contig of 16992 bp in length.
FEATURES                Location/Qualifiers
     source             1. .166668
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="12"
                        /clone="RP11-398O17"
BASE COUNT     58175 a  29138 c  28414 g  50819 t    122 others
ORIGIN

  Query Match             5.7%;  Score 56.4;  DB 2;  Length 166668;
  Best Local Similarity   58.1%;  Pred. No. 0.0069;
  Matches  137;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    2;

Qy        757 CCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGA 816
              |||||  | || ||  |  |  | | | | ||| |   || ||| ||  |
Db     111278 CCTTGAATATGTAGAGATAGAGAAGAAAGCAGTGGATACAGGTGGAGAGGTGGTCCAAAA 111219

Qy        817 TAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAG 876
              |||||  | || | |||||  |||| | || |||||| ||||||    |
Db     111218 ATGGGAAGATGTAGTTAAGAGGATACAAATTA--AGTTATGTAGGATGAACAAGTCTAGA 111161

Qy        877 AGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCAT 936
              | | |  |   | | ||  |    || |      |  | || | ||  |||||
Db     111160 GGTCTAAGATACAATATGAGGACTATAAATAATAAAATTGTACT-GTGTCTGTGATTCAT 111102

Qy        937 GCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              || ||||||||  |||||||||||||||||||||||  ||||||||||||||||
```

```
Db      111101 GCTAAATGAGTCGATTTTAGCTGCTCTTGCCACAAAACCAAAAAAAAAAAAAAAAAA 111046


RESULT 50
AC011004/c
LOCUS       AC011004               183778 bp    DNA     linear   HTG 11-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-110D15, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC011004
VERSION     AC011004.12  GI:13273359
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183778)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183778)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 11, 2001 this sequence version replaced gi:13270571.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 666
                 Center clone name: RP11-110D15
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 182108 bases at least Q40
                 Consensus quality: 183015 bases at least Q30
                 Consensus quality: 183336 bases at least Q20
                 Insert size: 185712; agarose-fp
                 Insert size: 183578; sum-of-contigs
                 Quality coverage:  9.3x in Q20 bases; agarose-fp
                 Quality coverage:  9.4x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    26915: contig of 26915 bp in length
            *    26916    27015: gap of unknown length
            *    27016    98873: contig of 71858 bp in length
            *    98874    98973: gap of unknown length
            *    98974   183778: contig of 84805 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183778
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-110D15"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .26915
                     /note="assembly_name:Contig8
                     clone_end:SP6"
     misc_feature    27016. .98873
                     /note="assembly_name:Contig9
                     clone_end:T7"
     misc_feature    98974. .183778
                     /note="assembly_name:Contig10"
BASE COUNT    50361 a  41935 c  41401 g  49880 t    201 others
ORIGIN

  Query Match             5.6%;  Score 56;  DB 2;  Length 183778;
  Best Local Similarity   64.4%;  Pred. No. 0.0087;
  Matches  116;  Conservative    0;  Mismatches   60;  Indels    4;  Gaps    2;

Qy        813 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 872
              |||| | || |   | |||| | ||||||||| ||  ||||  ||| |  ||| | || ||
Db      37638 AGGAAACGGGAGATGTAGGTGAAAGGACGCAGAGAGGCAGATATGCAAGATGCATCGCCC 37579

Qy        873 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 932
              ||||| | ||    |   |||||||| |  | ||| || | |  |    |  ||   ||  |
Db      37578 AGAGACCTAATGT-GCATGAGGACTCCTAGTCAACAACATTGA---ATTATACCGGGGAT 37523

Qy        933 TCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              |  ||| ||   |||   ||||||  |||||||||| |   |||||||||||||||||||
Db      37522 TTCTGCTAAGACAGTAGATTTTAGGAGCTCTTGCCATATTGAAAAAAAAAAAAAAAAAAAA 37463


Search completed: February  9, 2004, 14:16:33
Job time : 3928 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 13:03:52 ; Search time 2524 Seconds
(without alignments)
9552.314 Million cell updates/sec

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggcacgagccaggaactagg.....aaaaaaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	626.8	63.2	694	13	BX097129	BX097129 BX097129
c	2	409.8	41.3	414	14	H74303	H74303 yu56c11.s1
c	3	408	41.1	437	14	N54458	N54458 yv40b02.s1
	4	373.2	37.6	480	14	H73373	H73373 yu48f10.r1
	5	354	35.7	418	14	H58326	H58326 yr25c07.r1
	6	349	35.2	582	14	H74302	H74302 yu56c11.r1
	7	340.6	34.3	468	14	N76568	N76568 yv40b02.r1
	8	320.4	32.3	996	11	AK020081	AK020081 Mus muscu
c	9	286	28.8	405	9	AI242058	AI242058 qh81d07.x
c	10	284.2	28.6	296	14	H73374	H73374 yu48f10.s1
c	11	277.2	27.9	379	14	R02548	R02548 ye80a07.s1
c	12	274.2	27.6	450	9	AI438986	AI438986 tc84a08.x
	13	268.8	27.1	553	14	BY718164	BY718164 BY718164
	14	265.2	26.7	528	4	BX529784	Bx529784 RZPD Mus
c	15	261.2	26.3	384	14	H40263	H40263 yp59e11.s1
c	16	256.8	25.9	284	14	H58715	H58715 yr25c07.s1
c	17	235.2	23.7	376	14	T80187	T80187 yd85g03.s1
c	18	227.4	22.9	334	9	AA676653	AA676653 zj67e03.s
c	19	227	22.9	227	9	AI022453	AI022453 ow96h04.x
	20	220.2	22.2	358	14	R02666	R02666 ye80a07.r1
	21	218.4	22.0	1161	14	W41959	W41959 mc68b10.r1
	22	202	20.4	401	9	AA050499	AA050499 mj20f02.r
	23	192.2	19.4	324	14	T87491	T87491 yd85g03.r1
	24	182.2	18.4	467	14	W65713	W65713 me09g06.r1
	25	175.6	17.7	218	14	N86939	N86939 L1498F Huma
	26	174.6	17.6	364	13	BY212485	BY212485 BY212485
	27	154	15.5	537	12	BI347002	BI347002 376394 MA
	28	154	15.5	582	12	BI347003	BI347003 376395 MA
	29	153.8	15.5	443	14	CB051335	CB051335 NISC_gj23
c	30	153.8	15.5	540	14	CB051334	CB051334 NISC_gj23
	31	145.2	14.6	194	14	N88474	N88474 K3475F Huma
	32	143	14.4	382	10	BB842514	BB842514 BB842514
	33	132.6	13.4	370	10	BB843270	BB843270 BB843270
	34	93.2	9.4	420	10	BF443482	BF443482 260980 MA
	35	92.6	9.3	503	10	BF442092	BF442092 258654 MA
	36	79.8	8.0	208	9	AA031034	AA031034 mi47h05.r
c	37	57.8	5.8	483	9	AI187208	AI187208 qf28b11.x
c	38	52.8	5.3	466	9	AW794912	AW794912 RC6-UM001
c	39	52.6	5.3	493	2	HSM077193	Bx487006 Homo sapi
c	40	51.2	5.2	606	12	BI063151	BI063151 IL3-UT011
	41	50.8	5.1	443	13	BY560588	BY560588 BY560588
	42	49	4.9	925	29	CNS0091P	AL053013 Drosophil
	43	48	4.8	844	29	CNS0052P	AL056652 Drosophil
	44	47.4	4.8	471	9	AW795001	AW795001 RC6-UM001
c	45	47.2	4.8	974	13	BX339244	BX339244 BX339244
c	46	46.4	4.7	748	29	BZ772737	BZ772737 mcv51f02.
c	47	46.2	4.7	1201	13	BX356664	BX356664 BX356664
c	48	45.6	4.6	569	14	CB615918	CB615918 AMGNNUC:U
c	49	45.6	4.6	626	14	CD397038	CD397038 Gm_ck1771
	50	45.6	4.6	935	29	CNS006XK	AL066051 Drosophil
c	51	45.4	4.6	925	29	CNS0091P	AL053013 Drosophil
c	52	45.4	4.6	1201	13	BX381961	BX381961 BX381961
c	53	45	4.5	537	14	CA442247	CA442247 UI-H-DI0-
	54	44.8	4.5	739	29	AG187401	AG187401 Pan trogl
c	55	44.6	4.5	638	14	CB242205	CB242205 UI-CF-FN0
c	56	44.2	4.5	1100	29	CNS00FSE	AL070988 Drosophil
	57	43.4	4.4	660	13	BU928339	BU928339 AGENCOURT
c	58	43.2	4.4	420	9	AW333438	AW333438 S21G11 AG
c	59	43	4.3	449	9	AA186675	AA186675 zp71c08.s
	60	43	4.3	591	14	CB000366	CB000366 S345U_G08
	61	43	4.3	699	29	AG169292	AG169292 Pan trogl
c	62	43	4.3	940	13	BU185569	BU185569 AGENCOURT
c	63	42.6	4.3	349	28	AQ097268	AQ097268 HS_3038_B
c	64	42.6	4.3	524	28	AQ480601	AQ480601 RPCI-11-2
c	65	42.6	4.3	628	29	AG018533	AG018533 Homo sapi
c	66	42.6	4.3	638	29	AG018534	AG018534 Homo sapi
c	67	42.6	4.3	639	29	AG017811	AG017811 Homo sapi
c	68	42.6	4.3	641	29	AG017810	AG017810 Homo sapi
c	69	42.6	4.3	643	29	AG017808	AG017808 Homo sapi
c	70	42.4	4.3	699	14	CB046795	CB046795 NISC_gf06
c	71	42.2	4.3	385	9	AI478457	AI478457 tm44c09.x
c	72	42.2	4.3	469	13	BQ169652	BQ169652 WHE0961_B
	73	42.2	4.3	1036	9	AL550172	AL550172 AL550172
	74	42	4.2	543	9	AV597018	AV597018 AV597018
	75	42	4.2	549	28	AQ221080	AQ221080 HS_3223_B
	76	41.8	4.2	260	10	BF763246	BF763246 RC6-CS007
c	77	41.8	4.2	289	9	AW665077	AW665077 hi86h11.x

c	78	41.8	4.2	310	9	AW136417 AW136417 UI-H-BI1-
c	79	41.8	4.2	323	9	AI672594 AI672594 we55b06.x
c	80	41.8	4.2	384	9	AI492336 AI492336 ti23c01.x
c	81	41.8	4.2	386	9	AI521308 AI521308 ti04f12.x
c	82	41.8	4.2	387	9	AI337539 AI337539 qq02b09.x
c	83	41.8	4.2	387	9	AI540779 AI540779 tc87e02.x
c	84	41.8	4.2	399	9	AI400702 AI400702 tg93f07.x
c	85	41.8	4.2	402	9	AW572908 AW572908 hf17d02.x
c	86	41.8	4.2	403	9	AI091670 AI091670 oo23e05.x
c	87	41.8	4.2	409	9	AI184134 AI184134 qf30a08.x
c	88	41.8	4.2	417	9	AI924374 AI924374 wn60b11.x
c	89	41.8	4.2	430	10	BF591106 BF591106 7h53b05.x
c	90	41.8	4.2	431	9	AI458860 AI458860 tj55b09.x
c	91	41.8	4.2	432	12	BM766951 BM766951 K-EST0048
c	92	41.8	4.2	461	10	BF003058 BF003058 7g53d06.x
c	93	41.8	4.2	467	9	AI830572 AI830572 wj51d01.x
c	94	41.8	4.2	471	9	AI141026 AI141026 oy68h08.x
c	95	41.8	4.2	480	9	AI968111 AI968111 wu13c04.x
c	96	41.8	4.2	483	10	BF434111 BF434111 7o98h03.x
c	97	41.8	4.2	494	9	AI523802 AI523802 tg96c02.x
	98	41.8	4.2	694	14	CD366756 CD366756 UI-H-FT2-
c	99	41.8	4.2	717	10	BF109873 BF109873 7l70e11.x
c	100	41.8	4.2	900	13	BU839218 BU839218 AGENCOURT
c	101	41.8	4.2	918	13	BQ673943 BQ673943 AGENCOURT
c	102	41.8	4.2	923	13	BU172174 BU172174 AGENCOURT
c	103	41.8	4.2	971	11	CNS08ST4 BX026756 Single re
c	104	41.8	4.2	1474	13	BQ712584 BQ712584 AGENCOURT
	105	41.4	4.2	113	14	CB972450 CB972450 CAB30001_
c	106	41.4	4.2	128	14	CB972378 CB972378 CAB30001_
	107	41.4	4.2	351	14	CB250283 CB250283 EST0059 M
	108	41.4	4.2	560	13	BX118001 BX118001 BX118001
	109	41.4	4.2	1201	13	BX360624 BX360624 BX360624
c	110	41.2	4.2	303	14	CD393259 CD393259 Gm_ck1274
c	111	41.2	4.2	655	14	CD394654 CD394654 Gm_ck1439
	112	41.2	4.2	804	28	AQ898736 AQ898736 HS_3129_B
	113	41.2	4.2	931	13	BQ672426 BQ672426 AGENCOURT
c	114	41	4.1	839	13	BU852202 BU852202 AGENCOURT
	115	40.8	4.1	461	13	BQ216106 BQ216106 AGENCOURT
	116	40.8	4.1	498	28	AQ732239 AQ732239 HS_5533_B
	117	40.8	4.1	558	9	AU164477 AU164477
	118	40.8	4.1	691	28	BH693181 BH693181 BOHYD57TR
c	119	40.8	4.1	728	12	BI918323 BI918323 603182826
	120	40.8	4.1	979	13	BX415528 BX415528 BX415528
c	121	40.6	4.1	400	9	AA964525 AA964525 UI-R-C0-g
c	122	40.6	4.1	461	14	CD404072 CD404072 Gm_ck2688
	123	40.4	4.1	658	12	BI819899 BI819899 603035441
c	124	40.4	4.1	792	14	CB682466 CB682466 OSJNEf09N
c	125	40.4	4.1	917	13	BU590968 BU590968 AGENCOURT
c	126	40.4	4.1	1007	13	BX384114 BX384114 BX384114
	127	40.2	4.1	119	12	BI770253 BI770253 603055953
	128	40.2	4.1	312	9	AL373411 AL373411 MtBA57F07
	129	40.2	4.1	343	12	BM932065 BM932065 UI-E-EJ1-
c	130	40.2	4.1	372	28	AQ609683 AQ609683 HS_5079_A
	131	40.2	4.1	378	13	BQ252711 BQ252711 san79g01.
c	132	40.2	4.1	383	9	AA630611 AA630611 ac11d12.s
c	133	40.2	4.1	387	10	BF018663 BF018663 ux96e09.x
c	134	40.2	4.1	390	9	AW082128 AW082128 xb63a04.x
c	135	40.2	4.1	400	9	AI340960 AI340960 qo82d06.x
c	136	40.2	4.1	492	9	AA550826 AA550826 nj34c12.s
c	137	40.2	4.1	565	14	CD416197 CD416197 Gm_ck6495
c	138	40.2	4.1	608	9	AA148277 AA148277 zo44g12.s
c	139	40.2	4.1	616	14	CD392581 CD392581 Gm_ck1177
	140	40.2	4.1	685	14	CB240715 CB240715 UI-CF-FN0
c	141	40.2	4.1	828	11	CNS09NY2 BX067110 Single re
c	142	40.2	4.1	911	11	CNS09L4R BX063463 Single re
c	143	40.2	4.1	998	11	CNS09J93 BX061027 Single re
	144	40.2	4.1	1201	13	BX343351 BX343351 BX343351
	145	40.2	4.1	1431	29	BZ572186 BZ572186 msh2_2541
c	146	40	4.0	233	13	BQ149328 BQ149328 NF089D03F
c	147	40	4.0	528	9	AW264423 AW264423 xr03a11.x
	148	40	4.0	540	12	BI772784 BI772784 603058804
c	149	40	4.0	731	10	BG306497 BG306497 fm56g04.x
	150	40	4.0	740	13	BQ744040 BQ744040 WHE4111_A

ALIGNMENTS

RESULT 1
BX097129
LOCUS BX097129 694 bp mRNA linear EST 04-FEB-2003
DEFINITION BX097129 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGp998M04528 ; IMAGE:245163, mRNA sequence.
ACCESSION BX097129
VERSION BX097129.1 GI:27828666
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M., Radelof,U., Schneider,D. and Korn,B.
TITLE Human UnigeneSet - RZPD3
JOURNAL Unpublished
COMMENT Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGp998M04528.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human UnigeneSet - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer: M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES Location/Qualifiers
source 1. .694
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998M04528 ; IMAGE:245163"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen 1NFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 160 a 213 c 205 g 116 t
ORIGIN

Query Match 63.2%; Score 626.8; DB 13; Length 694;
Best Local Similarity 94.4%; Pred. No. 7.5e-79;
Matches 687; Conservative 0; Mismatches 2; Indels 39; Gaps 2;

```
Qy        223 CACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGG 282
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 CCCCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGG 64

Qy        283 TGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAG 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 TGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAG 124

Qy        343 ACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCA 402
```

```
Db        125 ACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCA 184

Qy        403 GGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCA 462
              ||||||||||||||||||||||||||||
Db        185 GGCTACAGATGCACTGGGAGCTGTGGTCCA---------------------------- 214

Qy        463 CTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCA 522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 --------GACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCA 266

Qy        523 GCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGAC 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGAC 326

Qy        583 CATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCT 642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 CATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCT 386

Qy        643 GGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAG 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 GGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAG 446

Qy        703 GTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGC 762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGC 506

Qy        763 CGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGA 822
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 CGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGA 566

Qy        823 AT-GGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCA 881
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 ATGGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACTGTCCAGAGAGCCA 626

Qy        882 AGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAA 941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        627 AGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAA 686

Qy        942 ATGAGTGT 949
              ||||||||
Db        687 ATGAGTGT 694
```

RESULT 2
H74303/c

```
LOCUS       H74303                 414 bp    mRNA    linear   EST 31-OCT-1995
DEFINITION  yu56c11.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:230132 3', mRNA sequence.
ACCESSION   H74303
VERSION     H74303.1  GI:1047714
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 414)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1114
            High quality sequence stops: 313
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1114   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 313.
FEATURES              Location/Qualifiers
     source           1. .414
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3781228"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:230132"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       75 a    127 c    116 g     95 t      1 others
ORIGIN
```

```
  Query Match           41.3%;  Score 409.8;  DB 14;  Length 414;
  Best Local Similarity  99.3%;  Pred. No. 2.6e-48;
  Matches  411;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        556 ATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCC 615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        414 ATGGGCAGGTCCACCTGCAGCAGAGACCNATCCACAGGCAGCCTGCCAACTTCTCCTTCC 355

Qy        616 TGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGC 675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        354 TGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGC 295

Qy        676 ACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTC 735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        294 ACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTC 235

Qy        736 CCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAG 795
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        234 CCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAG 175

Qy        796 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 855
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        174 AGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCA 115

Qy        856 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCA 915
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        114 TGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCA 55

Qy        916 CTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         54 CTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1
```

RESULT 3
N54458/c

```
LOCUS       N54458                 437 bp    mRNA    linear   EST 28-JAN-1997
DEFINITION  yv40b02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:245163 3', mRNA sequence.
ACCESSION   N54458
VERSION     N54458.1  GI:1195778
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 832   Std Error: 0.00
            Seq primer: m13 -40 forward
            High quality sequence stop: 375.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3794409"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:245163"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       78 a    132 c    119 g    106 t      2 others
ORIGIN

  Query Match             41.1%;  Score 408;  DB 14;  Length 437;
  Best Local Similarity   97.3%;  Pred. No. 4.5e-48;
  Matches  425;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy        537 AACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||    |  ||||
Db        437 AACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCAATNCCACAGCAG 378

Qy        597 CCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCA 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 CCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCA 318

Qy        657 AACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAG 716
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        317 AACAACGCCAATGTCCAACACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAG 258

Qy        717 ATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGC 776
              |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db        257 ATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCGTTCCTTGCCGCTCTACAGGAGC 198

Qy        777 ACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTC-AG 835
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        197 ACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCNAG 138

Qy        836 AGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGA 895
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        137 AGGACGCAAAGCAGCAGCCATGTAGAATGAACTGTCCAGAGAGCCAAGCACGGCAGAGGA 78

Qy        896 CTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTA 955
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 CTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTA 18

Qy        956 GCTGCTCTTGCCACAAA 972
              |||||||||||||||||
Db         17 GCTGCTCTTGCCACAAA 1


RESULT 4
H73373
LOCUS       H73373                   480 bp    mRNA    linear   EST 31-OCT-1995
DEFINITION  yu48f10.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:229387 5', mRNA sequence.
ACCESSION   H73373
VERSION     H73373.1  GI:1047623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1102
            High quality sequence stops: 306
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1102   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3780483"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:229387"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      100 a    149 c    130 g    100 t      1 others
ORIGIN

  Query Match             37.6%;  Score 373.2;  DB 14;  Length 480;
```

```
  Best Local Similarity   92.0%;  Pred. No. 3.3e-43;
  Matches  439;  Conservative    0;  Mismatches   29;  Indels    9;  Gaps    4;

Qy          12 GGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAGGCATGGGGCTC  71
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           4 GAAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAGGCATGGGGCTC  63

Qy          72 CCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAA 131
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          64 CCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAA 123

Qy         132 GAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGC 191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         124 GAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGC 183

Qy         192 TGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGT 251
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         184 TGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGT 243

Qy         252 GGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTC 311
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         244 GGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTC 303

Qy         312 AACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCT----CACCTACTTCTGCCGGGC 367
               ||||||||||||||||||||||||||||||||||||    |   |   |   | ||
Db         304 AACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTTCACCTNATTTCTTGCCGGGCG 363

Qy         368 GTCCTCCACCTCAGGTGCCCATGT-GGACAGTGCCAGGCT---ACAGATGCACTGGGAGC 423
               |||||||||||||||||||||||| |||||||||||||| |   |   |||||||||||
Db         364 TTCCTCCACCTCAGGTGCCCATGTGGGACAGTGCCAGGGTTACAGATGGCATTGGGGAGC 423

Qy         424 TGTGGTCCAAGCCAGTG-TCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGG 479
               |||||||||||||||| | |||| |||||||||| |||| | ||||||||||||||
Db         424 TGTGGTCCAAGCCAGTGTTTTGAGGTGCGGGCCAATTTCATTTTGCAGGACAGAGGG 480
```

```
RESULT 5
H58326
LOCUS       H58326                   418 bp    mRNA    linear   EST 05-OCT-1995
DEFINITION  yr25c07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:206316 5', mRNA sequence.
ACCESSION   H58326
VERSION     H58326.1  GI:1011158
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 912
            High quality sequence stops: 251
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 912   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 251.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3775447"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:206316"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       89 a    130 c    119 g     76 t      4 others
ORIGIN
```

```
  Query Match             35.7%;  Score 354;  DB 14;  Length 418;
  Best Local Similarity   96.1%;  Pred. No. 1.7e-40;
  Matches  393;  Conservative    0;  Mismatches   13;  Indels    3;  Gaps    3;

Qy         279 AAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGT 338
               || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           3 AGGTGGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGT  62

Qy         339 CCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGT 398
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          63 CCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGT 122

Qy         399 GCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAAC 458
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         123 GCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAAC 182

Qy         459 TTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCG 518
               ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db         183 TTCACTCTGCAGGACAGAGGGGCAGG-CCCAGGGTGGAGATGATCTGCCAGGCGTCCTCG 241

Qy         519 GGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAG 578
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         242 GGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTNCAGCAG 301

Qy         579 AGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCC-TGCCGAGCCAGACATCGGACTG 637
               |||||||||||||||  |||||||||||| |||||||||||||||||||||||||||||
Db         302 AGACCATGCCACAGGAGGCCTGCCAACTTGTCCTTCCTTGCCGAGCCAGACATCGGACTG 361

Qy         638 G-TTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCT 685
               | || |||||||||| |||||||||| |||||||||| |||| | |||
Db         362 GTTTTTGGTGCCAGGGTGCAAACAACGTCAATGTCCAGNACAGGGNCCT 410
```

```
RESULT 6
H74302
LOCUS       H74302                   582 bp    mRNA    linear   EST 31-OCT-1995
DEFINITION  yu56c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:230132 5', mRNA sequence.
ACCESSION   H74302
VERSION     H74302.1  GI:1047713
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
```

```
   MEDLINE   97044478
    PUBMED   8889549
 COMMENT     Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             Insert Size: 1114
             High quality sequence stops: 308
             Source: IMAGE Consortium, LLNL
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 1114   Std Error: 0.00
             Seq primer: M13RP1
             High quality sequence stop: 308.
FEATURES              Location/Qualifiers
     source           1. .582
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3781228"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:230132"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      126 a    149 c    157 g    130 t     20 others
ORIGIN

  Query Match             35.2%;  Score 349;  DB 14;  Length 582;
  Best Local Similarity   87.0%;  Pred. No. 7.4e-40;
  Matches  470;  Conservative    0;  Mismatches   53;  Indels   17;  Gaps    8;

Qy          10 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGA-GGCCCTACACCCACCGAGGCATGGGG 68
               ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db           1 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGANGGCCCTACACCCACCGAGGCATGGGG 60

Qy          69 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGC-TTCTCCAAGGCACGGGA 127
               ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          61 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTTCTCCAAGGCACGGGA 120

Qy         128 GGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGG 187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGG 180

Qy         188 CCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCT 247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 CCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCT 240

Qy         248 CTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTC 307
               |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         241 CTGTGGAACCAAGAACATCAAGGTTGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTC 300

Qy         308 CTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCT-GCTCACCTACTT--CTGCCG 364
               |||||||||||||||||||||||| |||||||||||| ||||||||| ||    ||||
Db         301 CTTCAACCTCAACGTCACACTCAAGTNCAGTCCAGACCTGGCTCACCTAATTTATGGCCG 360

Qy         365 GGCGTCCTCCACCTCAGGTGCCCATGTGG---ACAGTGCCAGGCTACAGATGCACTGGGA 421
               |||||||||||||| |||||||||||||   |     | || |||||||||| ||||
Db         361 GGCGTCCTCCACCTNAGGTGCCCATGTGGGACAGTTNCCAGGGTTACAGATGCATTGGGG 420

Qy         422 GCTGT---GGTCCAAGCCAGTGTCTGAGCT---GCGGGCCAACTTCACTCTGCAGGACAG 475
                   |   | |||||||||||||| |||| |   | || | |  | | | || |||||||
```

```
Db         421 AGGTTGTGGTTCCAAGCCAGTGTNTGAGGTTGCGGGGNCAATTTNAATTTTGNAGGACAG 480

Qy         476 AGGGGCAGGCCCCAGGGTGGAGATG---ATCTGCCAGGCGTCCTCGGGCAGCCCACCTAT 532
               |||||  | |  |||||||   ||      | |||||||  |  | ||| || |||  |||
Db         481 AGGGGNAAGGNCCAGGGTTNGGAGNTGATTTTGCCAGGNTTTTTAGGGNAGTCCAATTAT 540


RESULT 7
N76568
LOCUS       N76568                   468 bp    mRNA    linear   EST 28-JAN-1997
DEFINITION  yv40b02.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:245163 5', mRNA sequence.
ACCESSION   N76568
VERSION     N76568.1  GI:1239146
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 832   Std Error: 0.00
            Seq primer: reverse ET
            High quality sequence stop: 380.
FEATURES              Location/Qualifiers
     source           1. .468
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3794409"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:245163"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      105 a    156 c    126 g     80 t      1 others
ORIGIN

  Query Match             34.3%;  Score 340.6;  DB 14;  Length 468;
  Best Local Similarity   90.3%;  Pred. No. 1.2e-38;
  Matches  457;  Conservative    0;  Mismatches    5;  Indels   44;  Gaps    7;

Qy         218 CCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAA 277
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 CCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAA 61

Qy         278 GAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAG 337
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          62 GAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAG 121

Qy         338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         122 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 181

Qy         398 TGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAA 457
                |||||||||||||||||||||||||||||||||
Db         182 TGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCA------------------------- 216

Qy         458 CTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTC 517
                           |||||||||||||| ||||||||||||||||||||||||||||||||
Db         217 -------------GACAGAGGGGCAGG-CCCAGGGTGGAGATGATCTGCCAGGCGTCCTC 262

Qy         518 GGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCA 577
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         263 GGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCA 322

Qy         578 GAGACCATGCCACAGGCAGCCTGCCAAC-TTCTCCTTCCTGCCGAGCCAGACATCGGAC- 635
                ||||||||||||||   |||||||||| ||||||||||||||||||||||||||||||
Db         323 GAGACCATGCCACAGGAGGCCTGCCAACTTTCTCCTTCCTGCCGAGCCAGACATCGGACT 382

Qy         636 TGGTTCTGGTGCCAGGC-TGCAAACAACGCCAATG-TCCAGCACAGCGCCCTCACAGTGG 693
                |||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db         383 TGGTTCTGGTGCCAGGCTTGCAAACAACGCCAATGTTCCAGCACAGCGCCCTCACAGTGG 442

Qy         694 TGCCCCCAGG-TGGTGACCAGAAGAT 718
                |  ||||||| |||||| ||||||||
Db         443 TTGCCCCAGGTTGGTGANCAGAAGAT 468
```

RESULT 8
AK020081
LOCUS AK020081 996 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030468B19 product:hypothetical Immunoglobulin structure containing protein, full insert sequence.
ACCESSION AK020081
VERSION AK020081.1 GI:12860550
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 996)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES Location/Qualifiers
source 1. .996
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6030468B19"

```
                     /db_xref="MGI:1911589"
                     /db_xref="taxon:10090"
                     /clone="6030468B19"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
     CDS             33. .791
                     /note="unnamed protein product; hypothetical
                     Immunoglobulin structure containing protein (SCOP|48726,
                     evidence: SCOP)
                     putative"
                     /codon_start=1
                     /protein_id="BAB31985.1"
                     /db_xref="GI:12860551"
                     /db_xref="MGI:1924977"
                     /translation="MALLQLLLFAMLAACGFSEEQTEGITIAYKVLEVYPQSRRVLIT
                     CDAPEASQPITYSLLASRGILVAKKVVHDSVPASFNINITIKSSPDLLTYSCQATSNS
                     GTYGPSSRLQMYQELWAKPVSQLQADFVLRHGDSGPTVELSCLASSGSPPITYRLVGN
                     GGRVLAQQRPLHGKPANFSLPLSQTTGWFQCEAENDVGVDSSARIPLPRAEARAKLVT
                     TLAGELPLTPTCILAGSLVSIAVIASRMLSSTGL"
     polyA_signal    977. .982
                     /note="putative"
     polyA_site      996
                     /note="putative"
BASE COUNT       217 a    312 c    257 g    210 t
ORIGIN

  Query Match             32.3%;  Score 320.4;  DB 11;  Length 996;
  Best Local Similarity   68.9%;  Pred. No. 6e-36;
  Matches  460;  Conservative    0;  Mismatches  196;  Indels   12;  Gaps    1;

Qy         44 GCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGC 103
              |||| |  | ||| || |||||| ||||| |  |||| ||  ||| ||| ||||||||||
Db         14 GCCCCAACCTCACAGACGCATGGCGCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGC 73

Qy        104 CAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAA 163
              | |  |||||||  |||  | |   ||||  || |            ||||||||||||
Db         74 CTGTGGCTTCTCAGAGGAGCAGACAGAAGGCATCA------------CCATTGCCTACAA 121

Qy        164 AGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCC 223
              |||  ||||||||||  |||  || ||||  ||||||||  |||||||||  |||  ||  || |
Db        122 AGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCTGCGATGCCCCTGAGGC 181

Qy        224 ACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT 283
              | | |||||||||  ||  ||  |||   |  | |       |||  ||||||  || ||||||
Db        182 GTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGT 241

Qy        284 GGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGA 343
              |||  |   |  ||  |||||||||||||||  ||||  ||||  ||||||||||  |||||
Db        242 TGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGA 301

Qy        344 CCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAG 403
              ||||||||||||||  |||||  |||  ||||  |  |||  ||  ||  |||    |||   |||
Db        302 CCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCTATGGACCCAGCAGCAG 361

Qy        404 GCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCAC 463
              |||  ||||||  ||  |||  ||||||  |  ||||||||||||  ||||||  |||   |||||
Db        362 GCTCCAGATGTACCAGGAACTGTGGGCTAAGCCAGTGTCTCAGCTGCAGGCTGACTTCGT 421

Qy        464 TCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAG 523
              ||  |   |  |  |  |||||||  |||||  ||  ||||  |||  ||||  |||||
Db        422 CCTACGCCATGGGGACTCGGGCCCCACTGTGGAGCTGTCCTGCCTGGCATCCTCAGGCAG 481

Qy        524 CCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACC 583
              |||  ||  ||||||  ||  ||  |  |||||  |  |||||  ||  |     |||||  ||  ||
Db        482 CCCCCCCATCACCTACCGCTTGGTGGGGAATGGTGGGCGTGTTCTTGCACAGCAAAGGCC 541

Qy        584 ATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTG 643
              |   ||  |  | ||  |||||||||||||  |||  ||  |  ||||||  | |  ||||||
Db        542 ACTTCATGGAAAACCAGCCAACTTCTCCCTCCCGCTGTCCCAGACCACTGGTTGGTTCCA 601

Qy        644 GTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGG 703
```

```
              ||||  |  ||||  ||||  |  |  |||  |   |||  ||||  ||    |  ||||||  ||
Db        602 GTGCGAAGCTGAAAACGATGTCGGTGTGGACAGCAGTGCCCGCATCCCGCTGCCCCGAGC 661

Qy        704 TGGTGACC 711
                |  | ||
Db        662 AGAGGCCC 669


RESULT 9
AI242058/c
LOCUS       AI242058                405 bp    mRNA    linear   EST 01-DEC-1998
DEFINITION  qh81d07.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:1851085 3', mRNA sequence.
ACCESSION   AI242058
VERSION     AI242058.1  GI:3837455
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 522    Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1851085"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.  1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       95 a    113 c     81 g    116 t
ORIGIN

  Query Match             28.8%;  Score 286;  DB 9;  Length 405;
  Best Local Similarity   98.3%;  Pred. No. 6e-31;
  Matches  289;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        699 CCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCC 758
              | ||||||||||||||||||||||||||||  || ||||||||||||||||||||||||
Db        401 CAAGGTGGTGACCAGAAGATGGAGGACTGGAGGGTTCCCCTGGAGAGCCCCATCCTTGCC 342

Qy        759 TTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATA 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 TTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATA 282

Qy        819 GGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAG 878
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 GGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAG 222

Qy        879 CCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGC 938
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        221 CCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGC 162

Qy        939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        161 AAAATGAGTGTGTTTTAGCTGCTCTTGCCCCAAAAAAAAAAAAAAAAAAAAAAA 108


RESULT 10
H73374/c
LOCUS       H73374                   296 bp    mRNA    linear   EST 31-OCT-1995
DEFINITION  yu48f10.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:229387 3', mRNA sequence.
ACCESSION   H73374
VERSION     H73374.1  GI:1047624
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 296)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
   PUBMED   8889549
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1102
            High quality sequence stops: 182
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1102   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 182.
FEATURES             Location/Qualifiers
     source          1. .296
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3780483"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:229387"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       52 a     93 c     72 g     72 t      7 others
ORIGIN

  Query Match           28.6%;  Score 284.2;  DB 14;  Length 296;
  Best Local Similarity   96.6%;  Pred. No. 1.2e-30;
  Matches  286;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        658 ACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGA 717
              | |  ||||| ||||| |||||||||||| |||||||||||||||||||| |||||||||


Db        296 AAANGCCCAATNTCCAGNACAGCGCCCTCANAGTGGTGCCCCCAGGTGGTNACCAGAAGA 237

Qy        718 TGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCA 777
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        236 TGGAGGACTGGCAGGGTCCCCTGGAGANCCCCATCCTTGCCTTGCCGCTCTACAGGAGCA 177

Qy        778 CCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAG 837
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        176 CCCGCCGTCTGAGTNAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAG 117

Qy        838 GACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACT 897
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 GACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACT 57

Qy        898 GCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTT 953
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 GCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTT 1


RESULT 11
R02548/c
LOCUS       R02548                   379 bp    mRNA    linear   EST 31-MAR-1995
DEFINITION  ye80a07.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:124020 3', mRNA sequence.
ACCESSION   R02548
VERSION     R02548.1  GI:752284
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 846
            High quality sequence stops: 309 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 846   Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 309.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:476565"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:124020"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       71 a    121 c    103 g     83 t      1 others
ORIGIN
```

```
  Query Match           27.9%;  Score 277.2;  DB 14;  Length 379;
  Best Local Similarity  98.9%;  Pred. No. 1.1e-29;
  Matches  279;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        693 GTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATC 752
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 GTGGGCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATC 223

Qy        753 CTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTC 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        222 CTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTC 163

Qy        813 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 872
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        162 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACTGTCC 103

Qy        873 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 43

Qy        933 TCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
              |||||||||||||||||||||||||||||||||||||||||
Db         42 TCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 1


RESULT 12
AI438986/c
LOCUS       AI438986                 450 bp    mRNA    linear   EST 18-MAR-1999
DEFINITION  tc84a08.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2072822 3'
            similar to contains Alu repetitive element;contains element PTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AI438986
VERSION     AI438986.1  GI:4301160
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1792   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2072822"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CLL1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       85 a    144 c    109 g    112 t
ORIGIN

  Query Match           27.6%;  Score 274.2;  DB 9;  Length 450;
  Best Local Similarity  97.2%;  Pred. No. 2.6e-29;
  Matches  279;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        688 CAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 747
              || || |    | ||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CACTGCTTTGTCAAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 228

Qy        748 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 807
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        227 CCATCCTTGCCTTGCCGCTCTACAGGAGCGCCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 168

Qy        808 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 108

Qy        868 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 927
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 48

Qy        928 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
              |||||||||||||||||||||||||||||||||||||||||||||
Db         47 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 1


RESULT 13
BY718164
LOCUS       BY718164                 553 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY718164 RIKEN full-length enriched, 13 days embryo male testis Mus
            musculus cDNA clone 6030468B19 5', mRNA sequence.
ACCESSION   BY718164
VERSION     BY718164.1  GI:27131281
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
```

```
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..553
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6030468B19"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="13 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 13 days embryo
                     male testis"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 100.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT     118 a    184 c    137 g    114 t
ORIGIN

  Query Match            27.1%;  Score 268.8;  DB 14;  Length 553;
  Best Local Similarity  70.0%;  Pred. No. 1.3e-28;
  Matches  382;  Conservative    0;  Mismatches  152;  Indels   12;  Gaps    1;

Qy         44 GCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGC 103
Db         14 GCCCCAACCTCACAGACGCATGGCGCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGC 73

Qy        104 CAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAA 163
Db         74 CTGTGGCTTCTCAGAGGAGCAGACAGAAGGCATCA------------CCATTGCCTACAA 121

Qy        164 AGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCC 223
Db        122 AGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCTGCGATGCCCCTGAGGC 181

Qy        224 ACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT 283
Db        182 GTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGT 241

Qy        284 GGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGA 343
Db        242 TGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGA 301

Qy        344 CCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAG 403
Db        302 CCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCTATGGACCCAGCAGCAG 361

Qy        404 GCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCAC 463
Db        362 GCTCCAGATGTACCAGGAACTGTGGGCTAAGCCAGTGTCTCAGCTGCAGGCTGACTTCGT 421

Qy        464 TCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAG 523
Db        422 CCTACGCCATGGGGACTCGGGCCCCACTGTGGAGCTGTCCTGCCTGGCATCCTCAGGCAG 481

Qy        524 CCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACC 583
Db        482 CCCCCCCATCACCTACCGCTTGGTGGGGAATGGTGGGCGTGTTCTTGCACAGCAAAGGCC 541

Qy        584 ATGCCA 589
Db        542 ACTTCA 547


RESULT 14
BX529784
ID   BX529784   standard; RNA; EST; 528 BP.
XX
AC   BX529784;
XX
SV   BX529784.1
XX
DT   27-MAY-2003 (Rel. 75, Created)
DT   27-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   RZPD Mus musculus cDNA clone IMAGp998K041131 = IMAGE:476667 5' EST.
XX
KW   EST; expressed sequence tag.
XX
OS   Mus musculus (house mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-528
RA   Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,
RA   Korn B.;
RT   ;
RL   Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer
RL   Feld 580, D-69120 Heidelberg, Germany
XX
CC   RZPD; IMAGp998K041131.
CC   RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
CC   Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)
CC   http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981
CC   Contact: Ina Rolfs
```

```
CC   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC   Heubnerweg 6, D-14059 Berlin, Germany
CC   Tel: +49 30 32639 101
CC   Fax: +49 30 32639 111
CC   www.rzpd.de
CC   This clone is available royalty-free from RZPD;
CC   contact RZPD (clone@rzpd.de) for further information.
CC   Seq primer: T7, Primer sequence: TAATACGACTCACTATAGGG
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .528
FT                   /db_xref="taxon:10090"
FT                   /note="1st strand cDNA was primed with a Not I - oligo(dT)
FT                   primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTT
FT                   TTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2
FT                   13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru
FT                   Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was
FT                   ligated to EcoRI adaptors 5'-AATTCGGCACGAGG-3' and
FT                   5'-CCTCGTGCCG-3' (Pharmacia), digested with NotI and cloned
FT                   into the NotI and EcoRI sites of the pT7T3D-PacI vector.
FT                   Library went through one round of normalization, and was
FT                   constructed by Bento Soares and M.Fatima Bonaldo."
FT                   /organism="Mus musculus"
FT                   /clone="IMAGp998K041131"
FT                   /clone_lib="Soares 13.5-14.5dpc embryo NbME13.5-14.5"
FT                   /dev_stage="13.5-14.5dpc total fetus"
FT                   /lab_host="DH10B"
XX
SQ   Sequence 528 BP; 112 A; 177 C; 132 G; 107 T; 0 other;

  Query Match            26.7%;  Score 265.2;  DB 4;  Length 528;
  Best Local Similarity   71.6%;  Pred. No. 4.4e-28;
  Matches  348;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy        148 TCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCT 207
              || ||||||||||||||| |||||||||   ||| || ||||  |||||| |||||||
Db         42 TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCT 101

Qy        208 GCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCA 267
              ||  ||| ||  || |  | | |||||||||| || || |||    |  | |      |||
Db        102 GCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCC 161

Qy        268 AGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACAC 327
              |||||| || ||||| ||| |    |  || ||||||||||||||| ||||  ||||
Db        162 TGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCA 221

Qy        328 TCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCC 387
              |||||||||| |||||||||||||||||||| ||||| |||  |||| | ||| ||  ||
Db        222 TCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCT 281

Qy        388 ATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGC 447
              |||    |||   |||||| |||||| ||  ||| |||||| | |||||||||||| |||
Db        282 ATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTGTGGGCTAAGCCAGTGTCTCAGC 341

Qy        448 TGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCC 507
              ||| |||  ||||  || |   | | |  | |||||||  ||||||  ||  |||||
Db        342 TGCAGGCTGACTTCGTCCTACGCCATGGGGACTCGGGCCCCACTGTGGAGCTGTCCTGCC 401

Qy        508 AGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCC 567
              ||| ||||| |||||||| || |||||| || || || | ||||| | |||||  || |
Db        402 TGGCATCCTCAGGCAGCCCCCCCATCACCTACCGCTTGGTGGGGAATGGTGGGCGTGTTC 461

Qy        568 ACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGA 627
                  ||||| || |||   ||  |  | || ||||||||||| ||| || |  |||||
Db        462 TTGCACAGCAAAGGCCACTTCATGGAAAACCAGCCAACTTCTCCCTCCCGCTGTCCCAGA 521

Qy        628 CATCGG 633
              |  | |
Db        522 CCACTG 527


RESULT 15
```

```
H40263/c
LOCUS       H40263                  384 bp    mRNA    linear   EST 16-AUG-1995
DEFINITION  yp59e11.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:191756 3', mRNA sequence.
ACCESSION   H40263
VERSION     H40263.1  GI:916315
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1020
            High quality sequence stops: 297
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1020   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3761545"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:191756"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       70 a    122 c    111 g     77 t      4 others
ORIGIN

  Query Match            26.3%;  Score 261.2;  DB 14;  Length 384;
  Best Local Similarity   98.6%;  Pred. No. 1.8e-27;
  Matches  273;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy        693 GTGCCCCCAGGTGGTGA-CCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCAT 751
              |||  |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        278 GTGGGCCCAGGTGGTGACCCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCAT 219

Qy        752 CCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTT 811
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        218 CCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAANAGGAGTTTGGGGGGTT 159

Qy        812 CAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTC 871
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        158 CAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTC 99

Qy        872 CAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAG 931
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         98 CAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAG 39

Qy        932 TTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCA 968
              |||||||||||||||||||||||||||||||||||||
Db         38 TTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCA 2


RESULT 16
H58715/c
LOCUS       H58715                   284 bp    mRNA    linear   EST 05-OCT-1995
DEFINITION  yr25c07.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:206316 3', mRNA sequence.
ACCESSION   H58715
VERSION     H58715.1  GI:1011547
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 284)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 912
            High quality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 912   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 241.
FEATURES             Location/Qualifiers
     source          1. .284
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:3775447"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:206316"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       50 a    100 c     68 g     66 t
ORIGIN

  Query Match            25.9%;  Score 256.8;  DB 14;  Length 284;
  Best Local Similarity  99.2%;  Pred. No. 8.7e-27;
  Matches  258;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        693 GTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATC 752
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 GTGGGCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATC 201

Qy        753 CTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTC 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        200 CTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTC 141

Qy        813 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 81

Qy        873 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 932
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 21

Qy        933 TCATGCAAAATGAGTGTGTT 952
              ||||||||||||||||||||
Db         20 TCATGCAAAATGAGTGTGTT 1


RESULT 17
T80187/c
LOCUS       T80187                   376 bp    mRNA    linear   EST 15-MAR-1995
DEFINITION  yd85g03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:115060 3', mRNA sequence.
ACCESSION   T80187
VERSION     T80187.1  GI:698696
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1048
            High quality sequence stops: 330 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1048   Std Error: 0.00
            Seq primer: -21m13
            High quality sequence stop: 330.
FEATURES             Location/Qualifiers
     source          1. .376
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:470677"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:115060"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       65 a    117 c    114 g     78 t      2 others
ORIGIN

  Query Match            23.7%;  Score 235.2;  DB 14;  Length 376;
  Best Local Similarity  94.4%;  Pred. No. 8.2e-24;
  Matches  254;  Conservative    0;  Mismatches   14;  Indels    1;  Gaps    1;
```

```
Qy        686 CACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGT-CCCCTGGAGA 744
              | ||   | |||||  | |    ||||||||||||||||||||||| |||||||||| 
Db        269 CCCACGTGGGCCCCAGGTGGTGACCCAGAAGATGGAGGACTGGCAGGGTCCCCCTGGAGA 210

Qy        745 GCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTG 804
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        209 GCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAANAGGAGTTTG 150

Qy        805 GGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATG 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 GGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATG 90

Qy        865 AACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTA 924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 AACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTA 30

Qy        925 TTTGGAGTTCATGCAAAATGAGTGTGTTT 953
              |||||||||||||||||||||||||||||
Db         29 TTTGGAGTTCATGCAAAATGAGTGTGTTT 1
```

RESULT 18
AA676653/c
LOCUS AA676653 334 bp mRNA linear EST 19-DEC-1997
DEFINITION zj67e03.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:459964 3', mRNA sequence.
ACCESSION AA676653
VERSION AA676653.1 GI:2657175
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 334)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 100.
FEATURES Location/Qualifiers
source 1. .334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:459964"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 78 a 90 c 63 g 103 t

ORIGIN

Query Match 22.9%; Score 227.4; DB 9; Length 334;
Best Local Similarity 99.6%; Pred. No. 1.1e-22;
Matches 228; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        764 GCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAA 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 GCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAA 275

Qy        824 TGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAG 883
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 TGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAG 215

Qy        884 CACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAAT 943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 CACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAAT 155

Qy        944 GAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
              |||||||||||||||||||||||||||||||| |||||||||
Db        154 GAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAACAAAAAAAAA 106
```

RESULT 19
AI022453/c
LOCUS AI022453 227 bp mRNA linear EST 28-AUG-1998
DEFINITION ow96h04.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1654711 3', mRNA sequence.
ACCESSION AI022453
VERSION AI022453.1 GI:3237694
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 227)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 849 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 221.
FEATURES Location/Qualifiers
source 1. .227
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1654711"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 46 a 73 c 54 g 54 t
ORIGIN

Query Match 22.9%; Score 227; DB 9; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.5e-22;

```
  Matches  227;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        743 GAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTT 802
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTT 168

Qy        803 TGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAA 862
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 TGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAA 108

Qy        863 TGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCG 922
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCG 48

Qy        923 TATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              |||||||||||||||||||||||||||||||||||||||||||||||
Db         47 TATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1


RESULT 20
R02666
LOCUS       R02666                   358 bp    mRNA    linear   EST 31-MAR-1995
DEFINITION  ye80a07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:124020 5', mRNA sequence.
ACCESSION   R02666
VERSION     R02666.1  GI:752402
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 846
            High quality sequence stops: 309 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 846   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 309.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:476565"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:124020"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       74 a    111 c     99 g     68 t      6 others
ORIGIN
```

```
  Query Match          22.2%;  Score 220.2;  DB 14;  Length 358;
  Best Local Similarity  82.6%;  Pred. No. 1.1e-21;
  Matches  327;  Conservative    0;  Mismatches   23;  Indels   46;  Gaps    5;

Qy        282 GTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCA 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCA 60

Qy        342 GACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCC 401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCC 120

Qy        402 AGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTC 461
              ||||||||||||||||||||||||||||||
Db        121 AGGCTACAGATGCACTGGGAGCTGTGGTCCA---------------------------- 151

Qy        462 ACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGC 521
                          |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        152 ---------GACAGAGGGGCAGG-CCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGC 201

Qy        522 AGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAG-CAGAG 580
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| |||
Db        202 AGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGNCAGAG 261

Qy        581 ACCATGCCACAGGCAGCCTGCCAA---CTTCTCCTTCCTGCCGAGCCAGACATCGGACTG 637
              |||||||||||||   ||| | |       | | | ||||||||||||||||||| ||
Db        262 ACCATGCCACAGGAGNCCTTGCCAACTTNTTCCTTNCTTGCCGAGCCAGACATCGGATTG 321

Qy        638 GTTCTGG---TGCCAGGCTGCAAACAACGCCAATGT 670
              ||| | |     |  || |||||||||| || |||
Db        322 GTTNTTGGTTGCCAGGGTTGCAAACAACGNCATTGT 357


RESULT 21
W41959
LOCUS       W41959                   1161 bp    mRNA    linear   EST 21-MAY-1996
DEFINITION  mc68b10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:353659 5', mRNA sequence.
ACCESSION   W41959
VERSION     W41959.1  GI:1326461
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1161)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:225459
            Seq primer: ETPrimer
            High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:353659"
```

/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 185 a 251 c 470 g 208 t 47 others
ORIGIN

Query Match 22.0%; Score 218.4; DB 14; Length 1161;
Best Local Similarity 69.6%; Pred. No. 1.1e-21;
Matches 330; Conservative 0; Mismatches 131; Indels 13; Gaps 2;

```
Qy        66 GGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGG 125
             ||  |||| |  |||| ||  ||| ||| ||||||||||| |  ||||||  |||  | |
Db         6 GGCGTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGCCTGTGGCTTCTCAGAGGAGCAG 65

Qy       126 GAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAA 185
               ||||  || |           ||||||||||||||| ||||||||  ||| ||
Db        66 ACAGAAGGCATCA-----------CCATTGCCTACAAAGTACTGGAAGTTTATCCCCAA 113

Qy       186 GGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCC 245
               ||||   |||||| |||||||||  ||| ||  || |  | | |||||||||  || ||
Db       114 AGCCG-GAGGTGCTTATAACCTGCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCT 172

Qy       246 CTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCC 305
             |||   | | |      |||  |||||| || |||||  |||  |   |  || |||||||
Db       173 CTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCC 232

Qy       306 TCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGG 365
             |||||||||  ||||. ||||  |||||||||| ||||||||||||||||||| ||||| |
Db       233 TCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAG 292

Qy       366 GCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTG 425
             ||  |||| | |||  ||  || |||     |||    |||||| |||||| ||  ||| |||
Db       293 GCAACCTCGAACTCTGGCACCTATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTG 352

Qy       426 TGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGC 485
             |||  | |||||||||||| |||||| |||  |||||   ||  |   | | |   | |||
Db       353 TGGGCTAAGCCAGTGTCTCAGCTGCAGGCTGACTTCGTCCTACGCCATGGGGACTCGGGC 412

Qy       486 CCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAAC 539
             | ||  |||||| ||  |||||  ||| |||||  |||||||  | |||||  | |
Db       413 CGCACTGTGGAGCTGTCCTGCCTGGCATCCTCAGGCAGCCCCGCCATCACTACC 466
```

RESULT 22
AA050499
LOCUS AA050499 401 bp mRNA linear EST 09-SEP-1996
DEFINITION mj20f02.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone IMAGE:476667 5', mRNA sequence.
ACCESSION AA050499
VERSION AA050499.1 GI:1530170
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 401)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:287411
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 393.
FEATURES Location/Qualifiers
source 1. .401
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:476667"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT 87 a 130 c 101 g 83 t
ORIGIN

Query Match 20.4%; Score 202; DB 9; Length 401;
Best Local Similarity 72.4%; Pred. No. 3.6e-19;
Matches 262; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```
Qy       148 TCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCT 207
             || |||||||||||||||| |||||||||  |||  || ||||  |||||||  |||||||
Db        40 TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCT 99

Qy       208 GCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCA 267
             ||  |||  ||  ||  |  | |||||||| || ||  |||   |  | |   |||
Db       100 GCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCC 159

Qy       268 AGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACAC 327
             ||||||  ||  |||||  |||  |    |  ||  ||||||||||||||||  ||||  ||||
Db       160 TGGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCA 219

Qy       328 TCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCC 387
             |||||||||  |||||||||||||||||||  |||||  |||   ||||  |  |||  ||   ||
Db       220 TCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCT 279

Qy       388 ATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGC 447
             |||    |||   ||||||  ||||||  ||  |||  ||||||  |  ||||||||||||  |||
Db       280 ATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAACTGTGGGCTAAGCCAGTGTCTCAGC 339

Qy       448 TGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCC 507
             |||  |||  |||||    ||  |    |  |  |    |  ||||||   ||||||  ||  |||||
Db       340 TGCAGGCTGACTTCGTCCTACGCCATGGGGACTCGGGCCCCACTGTGGAGCTGTCCTGCC 399

Qy       508 AG 509
               |
```

```
Db        400 TG 401


RESULT 23
T87491
LOCUS       T87491                  324 bp    mRNA    linear   EST 17-MAR-1995
DEFINITION  yd85g03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:115060 5', mRNA sequence.
ACCESSION   T87491
VERSION     T87491.1  GI:715843
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1048
            High quality sequence stops: 179 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1048   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 179.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:470677"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:115060"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       70 a     99 c     77 g     78 t
ORIGIN

  Query Match             19.4%;  Score 192.2;  DB 14;  Length 324;
  Best Local Similarity   97.7%;  Pred. No. 9.4e-18;
  Matches  216;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy        131 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 134

Qy        191 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 194

Qy        251 TGGAACCAAGAACATCAAGGT-GGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC- 308
              ||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||
Db        195 TGGAACCAAGAACATCAAGGTGGGCCAAGAAGGTTGTGAAGACCCACGAGCCGGCCTCCT 254

Qy        309 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCT 349
              |||||||| |||||||||||||||||||||||||||||| |
Db        255 TTCAACCTTAACGTCACACTCAAGTCCAGTCCAGACCTGGT 295


RESULT 24
W65713
LOCUS       W65713                  467 bp    mRNA    linear   EST 11-JUN-1996
DEFINITION  me09g06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:387034 5', mRNA sequence.
ACCESSION   W65713
VERSION     W65713.1  GI:1373922
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:238866
            Seq primer: ETPrimer
            High quality sequence stop: 345.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:387034"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       97 a    159 c    112 g     99 t
ORIGIN

  Query Match             18.4%;  Score 182.2;  DB 14;  Length 467;
  Best Local Similarity   68.0%;  Pred. No. 2e-16;
  Matches  319;  Conservative    0;  Mismatches  133;  Indels   17;  Gaps    4;

Qy         66 GGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGG 125
              || |||| |  |||| ||  ||| ||| ||||||||||| |  | |     | |   |
Db         14 GGCCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGCCTGTGGTTCTCAGAGGAGCAGA 73
```

```
Qy        126 GAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAA 185
              || | | |              || ||||||||||||||||| ||||||||||  ||| ||
Db         74 CAGAAGGCA------------TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAA 120

Qy        186 GGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCC 245
              ||||   |||||| |||||||||  ||| ||  || |  | | |||||||| || ||
Db        121 AGCCGGA-GGTGCTTATAACCTGCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCT 179

Qy        246 CTCTGTGGAA--CCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGG 303
              |||    |  |  | | |     |  ||||| || ||||| ||| |   |  || |||||
Db        180 CTCCTGGCTAGCCGATGGTATCCTGCGTGGCAAAAAAGGTTGTGCATGACTCCGTGCCGG 239

Qy        304 CCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCC 363
              |||||||||| ||||  |||| |||||||| |||||||||||||||||| |||||
Db        240 CCTCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGACCTGCTCACCTACTCCTGCC 299

Qy        364 GGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGC 423
              |||  |||| | ||| ||  || |||    |||   |||||| |||||| ||  ||| |
Db        300 AGGCAACCTCGAACTCTGGCACCTATGGACCCAGCAGCAGGCTCCAGATGTACCAGGAAC 359

Qy        424 TGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAG 483
              ||||| | ||| |||||||| |||||| |||  |||||   || |   | | |  | |
Db        360 TGTGGGCTAAG-CAGTGTCTCAGCTGCAGGCTGACTTCGTCCTACGCCATGGGGACTCGG 418

Qy        484 GCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTAT 532
              ||||||  |||||| ||  ||||| ||| ||||| |||||||| || ||
Db        419 GCCCCACTGTGGAGCTGTCCTGCCTGGCATCCTCAGGCAGCCCCCCCAT 467
```

RESULT 25
N86939
LOCUS N86939 218 bp mRNA linear EST 01-APR-1996
DEFINITION L1498F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone L1498 5' similar to EST(YU56C11.S1), mRNA sequence.
ACCESSION N86939
VERSION N86939.1 GI:1440141
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 218)
AUTHORS Liew,C.C.
TITLE cDNAs from fetal heart (1996)
JOURNAL Unpublished
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GAAATTAACCCTCACTAAAGGG.
FEATURES Location/Qualifiers
source 1. .218
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L1498"
/lab_host="E. coli XL1-Blue"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."
BASE COUNT 50 a 48 c 71 g 49 t
ORIGIN

Query Match 17.7%; Score 175.6; DB 14; Length 218;
Best Local Similarity 97.1%; Pred. No. 2.4e-15;
Matches 200; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
Qy        750 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAG-TTTGGGGG 808
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTTGGGGG 60

Qy        809 GTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACC 868
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db         61 GTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGATTGAATC 120

Qy        869 GTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTG 928
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        121 GTCCAGAGAGCCAAGCACGGCAGAGGACTTCAGGCCATCAGCGTGCACTGTTCGTATTTG 180

Qy        929 GAGTTCATGCAAAATGAGTGTGTTTT 954
              | |||||||| ||||||||||||||
Db        181 GGGTTCATGC-AAATGAGTGTGTTTT 205
```

RESULT 26
BY212485
LOCUS BY212485 364 bp mRNA linear EST 10-DEC-2002
DEFINITION BY212485 RIKEN full-length enriched, activated spleen Mus musculus cDNA clone F830016F14 5', mRNA sequence.
ACCESSION BY212485
VERSION BY212485.1 GI:26393130
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 364)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.

, Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K. , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .364
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F830016F14"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched, activated spleen"
BASE COUNT       84 a    123 c     82 g     74 t      1 others
ORIGIN

  Query Match             17.6%;  Score 174.6;  DB 13;  Length 364;
  Best Local Similarity   69.6%;  Pred. No. 2.6e-15;
  Matches  256;  Conservative    0;  Mismatches  100;  Indels   12;  Gaps    1;

Qy          44 GCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGC 103
               |||| |  | ||| || |||||| ||||| |  |||| ||  ||| ||| ||||||||||
Db           9 GCCCCAACCTCACAGACGCATGGCGCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGC 68

Qy         104 CAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAA 163
               | |  |||||| ||| | |  |||| || |           ||||||||||||
Db          69 CTGTGGCTTCTCAGAGGAGCAGACAGAAGGCATCA-----------CCATTGCCTACAA 116

Qy         164 AGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCC 223
               ||| ||||||||||  ||| || ||||  |||||||| |||||||| ||| ||  || |
Db         117 AGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCTGCGATGCCCCTGAGGC 176

Qy         224 ACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT 283
                 | | |||||||||  || ||  |||    | | |       |||  |||||| || |||||
Db         177 GTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGT 236

Qy         284 GGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGA 343
                 |||  |   || ||||||||||||||||  ||||  ||||  ||||||||||  |||||
Db         237 TGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGA 296

Qy         344 CCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAG 403
               ||||||||||||||  |||||  |||  ||||  | ||| ||  ||  |||    |||    ||
Db         297 CCTGCTCACCTACTCCTGCCAGGCAACCTCGAACTCTGGCACCTATGGACCCAGCAGNAG 356

Qy         404 GCTACAGA 411
               | | ||||
Db         357 GGTCCAGA 364


RESULT 27
BI347002
```

```
LOCUS       BI347002                537 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  376394 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI347002
VERSION     BI347002.1  GI:15040291
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
  TITLE     Porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
  JOURNAL   Mamm. Genome 13 (8), 475-478 (2002)
  MEDLINE   22213789
   PUBMED   12226715
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 130  row: N  column: 7
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2PIG"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      110 a    182 c    154 g     91 t
ORIGIN

  Query Match             15.5%;  Score 154;  DB 12;  Length 537;
  Best Local Similarity   75.2%;  Pred. No. 1.7e-12;
  Matches  206;  Conservative    0;  Mismatches   65;  Indels    3;  Gaps    1;

Qy         433 AGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG---GGCAGGCCCCA 489
               |||| |||||  |||||| ||  |||||||  |||| |||| ||||    | | |||||
Db         132 AGCCCGTGTCCCAGCTGCTGGTTAACTTCATCCTGCTGGACGGAGGCCCGTCGGGCCCTT 191

Qy         490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
               ||||| || ||   ||| ||| ||||||||||||||||| ||||||  |||| ||  |||
Db         192 GGGTGAAGGTGTCGTGCTGGGCATCCTCGGGCAGCCCACCCATCACCTACAGTCTTGTCG 251

Qy         550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
               ||| |||  |       | || || ||||||  ||     | ||| | |  ||||||||||||
Db         252 GGAGGGACAGTGGCACCTACATGAAGCAGACACAGAACTACAAGGAAGCTGCCAACTTCT 311

Qy         610 CCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATG 669
               ||||||  ||   | ||||||  |||  |||| || ||||||||||||||||||  ||  |
Db         312 CCTTCCCGCTCACCCAGACGTCGTGCTGGCTCAGGTGCCAGGCTGCAAACAACATCAGCG 371

Qy         670 TCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGG 703
                |||||||| |||||||| ||||||||||||
Db         372 CCCAGCACAGTGCCCTCACACTGGTGCCCCCAGG 405


RESULT 28
BI347003
```

```
LOCUS       BI347003                582 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  376395 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI347003
VERSION     BI347003.1  GI:15040292
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
  TITLE     Porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
  JOURNAL   Mamm. Genome 13 (8), 475-478 (2002)
  MEDLINE   22213789
   PUBMED   12226715
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 130  row: N  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2PIG"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      122 a    191 c    166 g    103 t
ORIGIN

  Query Match             15.5%;  Score 154;  DB 12;  Length 582;
  Best Local Similarity   75.2%;  Pred. No. 1.7e-12;
  Matches  206;  Conservative    0;  Mismatches   65;  Indels    3;  Gaps    1;

Qy        433 AGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG---GGCAGGCCCCA 489
              |||| |||||  ||||||  ||  |||||||  ||||  ||||  ||||    | | |||||
Db        132 AGCCCGTGTCCCAGCTGCTGGTTAACTTCATCCTGCTGGACGGAGGCCCGTCGGGCCCTT 191

Qy        490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
              |||||  ||  ||   |||  ||| ||||||||||||||||  ||||||  ||||  ||   |||
Db        192 GGGTGAAGGTGTCGTGCTGGGCATCCTCGGGCAGCCCACCCATCACCTACAGTCTTGTCG 251

Qy        550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
              |||  |||   |      | ||  ||  ||||||  ||    | |||  | |  ||||||||||||
Db        252 GGAGGGACAGTGGCACCTACATGAAGCAGACACAGAACTACAAGGAAGCTGCCAACTTCT 311

Qy        610 CCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATG 669
              ||||||  ||   | ||||||  |||   ||||  ||  ||||||||||||||||||  ||   |
Db        312 CCTTCCCGCTCACCCAGACGTCGTGCTGGCTCAGGTGCCAGGCTGCAAACAACATCAGCG 371

Qy        670 TCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGG 703
                |||||||||  |||||||||  ||||||||||||
Db        372 CCCAGCACAGTGCCCTCACACTGGTGCCCCCAGG 405


RESULT 29
CB051335
LOCUS       CB051335                443 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  NISC_gj23h11.y1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3289604
            5', mRNA sequence.
ACCESSION   CB051335
VERSION     CB051335.1  GI:27789622
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM8055  row: P  column: 21
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3289604"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Pr28"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      129 a    112 c    111 g     91 t
ORIGIN

  Query Match             15.5%;  Score 153.8;  DB 14;  Length 443;
  Best Local Similarity   98.7%;  Pred. No. 2e-12;
  Matches  155;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        833 CAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGA 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGA 309

Qy        893 GGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTT 952
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        310 GGACTGCAGGCCATCAGCATGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTT 369

Qy        953 TTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAA 989
              |||||||||||||||| ||||||||||||||||||
Db        370 TTAGCTGCTCTTGCCACAGAAAAAAAAAAAAAAAAAA 406


RESULT 30
CB051334/c
LOCUS       CB051334                540 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  NISC_gj23h11.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3289604
            3', mRNA sequence.
ACCESSION   CB051334
VERSION     CB051334.1  GI:27789621
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM8055  row: P  column: 21
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3289604"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Pr28"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT      126 a    113 c    128 g    173 t
ORIGIN

  Query Match             15.5%;  Score 153.8;  DB 14;  Length 540;
  Best Local Similarity   98.7%;  Pred. No. 1.8e-12;
  Matches  155;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        833 CAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGA 892
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGA 246

Qy        893 GGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTT 952
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        245 GGACTGCAGGCCATCAGCATGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTT 186

Qy        953 TTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 989
              |||||||||||||||| ||||||||||||||||||||
Db        185 TTAGCTGCTCTTGCCCCAAAAAAAAAAAAAAAAAAAA 149


RESULT 31
N88474
LOCUS       N88474                  194 bp    mRNA    linear   EST 02-APR-1996
DEFINITION  K3475F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
            clone K3475 5', mRNA sequence.
ACCESSION   N88474
VERSION     N88474.1  GI:1441676
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194)
  AUTHORS   Liew,C.C.
  TITLE     cDNAs from fetal heart (1996)
  JOURNAL   Unpublished
COMMENT     Contact: Liew CC
            Brigham and Women's Hospital
            Harvard Medical School
            75 Francis St. Boston, MA 02115, USA
            Tel: 6177328915
            Fax: 6179750995
            Email: cliew@rics.bwh.harvard.edu
            Seq primer: GAAATTAACCCTCACTAAAGGG.
FEATURES             Location/Qualifiers
     source          1. .194
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="K3475"
                     /lab_host="E. coli XL1-Blue"
                     /clone_lib="Human fetal heart, Lambda ZAP Express"
                     /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
                     XhoI; mRNA was purified from human fetal hearts (8-10
                     weeks).  cDNA was synthesized using a XhoI-Oligo dT
                     adaptor-primer.  EcoRI adaptors were ligated, followed by
                     digestion with XhoI, for directional cloning into
                     predigested lambda ZAP Express."
BASE COUNT      51 a     43 c     63 g     37 t
ORIGIN

  Query Match             14.6%;  Score 145.2;  DB 14;  Length 194;
  Best Local Similarity   96.5%;  Pred. No. 4.6e-11;
  Matches  191;  Conservative    0;  Mismatches    3;  Indels    4;  Gaps    4;

Qy        750 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 809
              |||||||||||||||||||||||||||||||| |||||||||||||||||||| |||||
Db          1 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCCCCGTCTGAGTGAAGAGGAGTTT-GGGGG 59

Qy        810 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCG 869
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db         60 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATCTAGAATGAATCG 119

Qy        870 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 929
              | |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db        120 T-CAGAGAGCCAAGCACGGCAGAGGACTGCAGG-CATCAGCGTGCACTG-TCGTATTTGG 176

Qy        930 AGTTCATGCAAAATGAGT 947
              ||||||||||||||||||
Db        177 AGTTCATGCAAAATGAGT 194


RESULT 32
BB842514
LOCUS       BB842514                382 bp    mRNA    linear   EST 21-NOV-2001
DEFINITION  BB842514 RIKEN full-length enriched, 6 days neonate spleen Mus
            musculus cDNA clone F430005P21 5', mRNA sequence.
ACCESSION   BB842514
VERSION     BB842514.1  GI:17043245
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
```

```
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F430005P21"
                     /tissue_type="spleen"
                     /dev_stage="6 days neonate"
                     /clone_lib="RIKEN full-length enriched, 6 days neonate
                     spleen"
BASE COUNT       87 a    122 c     99 g     74 t
ORIGIN

  Query Match           14.4%;  Score 143;  DB 10;  Length 382;
  Best Local Similarity  69.8%;  Pred. No. 7e-11;
  Matches  213;  Conservative    0;  Mismatches   80;  Indels   12;  Gaps    1;

Qy         44 GCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGC 103
              |||| |  | ||| || |||||| ||||| |  |||| ||  ||| ||| ||||||||||
Db         90 GCCCCAACCTCACAGACGCATGGCGCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGC 149

Qy        104 CAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAA 163
              | |  |||||||  |||  | |   ||||  || |            ||||||||||||
Db        150 CTGTGGCTTCTCAGAGGAGCAGACAGAAGGCATCA-----------CCATTGCCTACAA 197

Qy        164 AGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCC 223
              ||| ||||||||||  ||| || ||||  |||||||  ||||||||  |||  ||  || |
Db        198 AGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCTGCGATGCCCCTGAGGC 257

Qy        224 ACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT 283
              | | |||||||||| || || |||    | | |       |||  ||||||  || |||||
Db        258 GTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGT 317

Qy        284 GGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGA 343
              |||  |   |  || ||||||||||||||||  ||||  ||||  ||||||||| |||||
Db        318 TGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCATCAAGTCCAGCCCAGA 377

Qy        344 CCTGC 348
              |||||
Db        378 CCTGC 382


RESULT 33
BB843270
LOCUS       BB843270                370 bp    mRNA    linear   EST 26-NOV-2001
DEFINITION  BB843270 RIKEN full-length enriched, 6 days neonate spleen Mus
            musculus cDNA clone F430010H21 5', mRNA sequence.
ACCESSION   BB843270
VERSION     BB843270.1  GI:17081637
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F430010H21"
                     /tissue_type="spleen"
                     /dev_stage="6 days neonate"
                     /clone_lib="RIKEN full-length enriched, 6 days neonate
                     spleen"
BASE COUNT       85 a    116 c     96 g     73 t
ORIGIN

  Query Match           13.4%;  Score 132.6;  DB 10;  Length 370;
  Best Local Similarity  68.9%;  Pred. No. 2e-09;
  Matches  202;  Conservative    0;  Mismatches   79;  Indels   12;  Gaps    1;

Qy         44 GCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGC 103
              |||| |  | ||| || |||||| ||||| |  |||| ||  ||| ||| ||||||||||
Db         90 GCCCCAACCTCACAGACGCATGGCGCTCCTTCAGCTGCTCCTCTTTGCCATGCTGGCTGC 149
```

```
Qy        104 CAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAA 163
              | |  ||||||  |||  | |  ||||  || |           ||||||||||||
Db        150 CTGTGGCTTCTCAGAGGAGCAGACAGAAGGCATCA-----------CCATTGCCTACAA 197

Qy        164 AGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCC 223
              ||| ||||||||||  ||| || ||||  |||||||  |||||||||  |||  ||  || |
Db        198 AGTACTGGAAGTTTATCCCCAAAGCCGGAGGGTGCTTATAACCTGCGATGCCCCTGAGGC 257

Qy        224 ACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGT 283
              | | |||||||||| || || |||   | | |       |||  ||||||  || |||||
Db        258 GTCCCAGCCCATCACATACTCTCTCCTGGCTAGCCGAGGTATCCTGGTGGCAAAAAAGGT 317

Qy        284 GGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCA 336
              |||  |   |  || ||||||||||||||||  ||||  ||||  |||||||||
Db        318 TGTGCATGACTCCGTGCCGGCCTCCTTCAACATCAATATCACCATCAAGTCCA 370
```

RESULT 34
BF443482

```
LOCUS       BF443482                 420 bp    mRNA    linear   EST 01-DEC-2000
DEFINITION  260980 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF443482
VERSION     BF443482.1  GI:11503574
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 420)
  AUTHORS   Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
  TITLE     Porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
  JOURNAL   Mamm. Genome 13 (8), 475-478 (2002)
  MEDLINE   22213789
   PUBMED   12226715
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 93  row: K  column: 24
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .420
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2PIG"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT       99 a    109 c    130 g     82 t
ORIGIN

  Query Match             9.4%;  Score 93.2;  DB 10;  Length 420;
  Best Local Similarity   71.1%;  Pred. No. 0.00064;
  Matches  138;  Conservative    0;  Mismatches   53;  Indels    3;  Gaps    1;

Qy        428 GTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG---GGCAGG 484
              | |  |||| |||||  ||||||  ||  |||||||  |||| |||| ||||    | | ||
Db        223 GCCAGAGCCCGTGTCCCAGCTGCTGGTTAACTTCATCCTGCTGGACGGAGGCCCGTCGGG 282

Qy        485 CCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCT 544
              |||  |||||  ||  ||    |||  ||| ||||||||||||||||||  ||||||  |||| ||
Db        283 CCCTTGGGTGAAGGTGTCGTGCTGGGCATCCTCGGGCAGCCCACCCATCACCTACAGTCT 342

Qy        545 GATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAA 604
                ||||||  |||   |     | ||  ||  ||||||  ||     |  |||  | |  |||||||
Db        343 TGTCGGGAGGGACAGTGGCACCTACATGAAGCAGACACAGAACTACAAGGAAGCTGCCAA 402

Qy        605 CTTCTCCTTCCTGC 618
              ||||||||||| ||
Db        403 CTTCTCCTTCCCGC 416
```

RESULT 35
BF442092

```
LOCUS       BF442092                 503 bp    mRNA    linear   EST 01-DEC-2000
DEFINITION  258654 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF442092
VERSION     BF442092.1  GI:11502184
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,J., White,J.,
            Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
            ,J. and Keele,J.W.
  TITLE     Porcine gene discovery by normalized cDNA-library sequencing and
            EST cluster assembly
  JOURNAL   Mamm. Genome 13 (8), 475-478 (2002)
  MEDLINE   22213789
   PUBMED   12226715
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 88  row: H  column: 8
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 2PIG"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      104 a    160 c    160 g     79 t
ORIGIN

  Query Match             9.3%;  Score 92.6;  DB 10;  Length 503;
  Best Local Similarity   70.8%;  Pred. No. 0.00072;
  Matches  138;  Conservative    0;  Mismatches   54;  Indels    3;  Gaps    1;

Qy        433 AGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG---GGCAGGCCCCA 489
              |||| |||||  |||||| ||  |||||||  |||| |||| ||||    | | |||||
Db        309 AGCCCGTGTCCCAGCTGCTGGTTAACTTCATCCTGCTGGACGGAGGCCCGTCGGGCCCTT 368

Qy        490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
              |||| || ||   |||  ||| ||||||||||||||||||  ||||||  |||| ||  |||
Db        369 GGGTGAAGGTGTCGTGCTGGGCATCCTCGGGCAGCCCACCCATCACCTACAGTCTTGTCG 428
```

```
Qy        550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
              ||| ||| | | || || |||||| ||  | ||| | ||||||||||||
Db        429 GGAGGGACAGTGGCACCTACATGAAGCAGACACAGAACTACAAAGAAGCTGCCAACTTCT 488

Qy        610 CCTTCCTGCCGAGCC 624
              |||||| ||  | ||
Db        489 CCTTCCCGCTCACCC 503


RESULT 36
AA031034
LOCUS       AA031034                208 bp    mRNA    linear   EST 21-AUG-1996
DEFINITION  mi47h05.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:466713 5', mRNA sequence.
ACCESSION   AA031034
VERSION     AA031034.1  GI:1501031
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:280529
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 193.
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:466713"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       47 a     61 c     51 g     49 t
ORIGIN

  Query Match             8.0%;  Score 79.8;  DB 9;  Length 208;
  Best Local Similarity   65.3%;  Pred. No. 0.067;
  Matches  139;  Conservative    0;  Mismatches   62;  Indels   12;  Gaps    1;

Qy         77 GCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAAT 136
              |||| ||  ||| ||| |||||||||| |    |||||  |||  | |   ||||  |
Db          7 GCTGCTCCTCTTTGCCATGCTGGCTGCCTGTGCGTTCTCAGAGGAGCAGACAGAAGGCA- 65

Qy        137 TACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGT 196
                  || |||||||||||||||| |||||||||  ||| || ||||  ||
Db         66 ----------TCACCATTGCCTACAAAGTACTGGAAGTTTATCCCCAAAGCCGGAGGTA 114

Qy        197 GCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAAC 256
              ||| |||||||||  ||| ||  || |  | | |||||||||  || || |||   |  |
Db        115 GCTTATAACCTGCGATGCCCCTGAGGCGTCCCAGCCCATCACATACTCTCTCCTGGCTAG 174

Qy        257 CAAGAACATCAAGGTGGCCAAGAAGGTGGTGAA 289
              |      |||  |||||| || ||||| ||| |
Db        175 CCGAGGTATCCTGGTGGCAAAAAAGGTTGTGCA 207


RESULT 37
AI187208/c
LOCUS       AI187208                483 bp    mRNA    linear   EST 10-NOV-1998
DEFINITION  qf28b11.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1751325
            3' similar to contains MER27.t1 MER32 repetitive element ;, mRNA
            sequence.
ACCESSION   AI187208
VERSION     AI187208.1  GI:3737846
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 483)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 678   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 381.
FEATURES             Location/Qualifiers
     source          1. .483
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1751325"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      133 a    103 c     78 g    169 t
ORIGIN

  Query Match             5.8%;  Score 57.8;  DB 9;  Length 483;
  Best Local Similarity   67.9%;  Pred. No. 56;
  Matches  112;  Conservative    0;  Mismatches   47;  Indels    6;  Gaps    2;

Qy        829 AGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC--CAAGCAC 886
```

```
Db        284 AGGTCAGAGCATACAAAGTAGCAGGTATGTAGGATGAACAAGGCTAGAATCTACTGTATA 225

Qy        887 GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAG 946

Db        224 ACATGAGGACTGTAGGTAATAAAATTGTACT----GTATGTGGAGTTCATGCTAAATGAG 169

Qy        947 TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 991

Db        168 TAGATTTTAGCTGCTCTTGCCACAAGAACAAAAAATCTAAGGGAA 124


RESULT 38
AW794912/c
LOCUS       AW794912                466 bp    mRNA    linear   EST 16-MAY-2000
DEFINITION  RC6-UM0015-210200-011-D07 UM0015 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW794912
VERSION     AW794912.1  GI:7846691
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC6-UM0015-210
            200-011-D07&t3=2000-02-21&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 14
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UM0015"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      113 a     98 c     84 g    171 t
ORIGIN

  Query Match             5.3%;  Score 52.8;  DB 9;  Length 466;
  Best Local Similarity   58.5%;  Pred. No. 2.9e+02;
  Matches  110;  Conservative    0;  Mismatches   77;  Indels    1;  Gaps    1;

Qy        804 GGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAAT 863

Db        461 GGGTGGAAAGAAAATGGGGAGACGAAAGTCAGAGGATGTAAAGTAGCAGACACATAGAAT 402

Qy        864 GAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGT 923

Db        401 GATAAAGTCTAGCTTTCTTATGTGCAACATGAGGACTACAGATAATAATATAGTGCT-GT 343

Qy        924 ATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAA 983

Db        342 ATCTGGGATTCATGATAAATGAGTAGATTTTAGTCGGTCCAGCCACAAAAGCAAAAATAA 283

Qy        984 AAAAAAAA 991

Db        282 AAACAAAA 275


RESULT 39
HSM077193/c
ID   HSM077193  standard; RNA; EST; 493 BP.
XX
AC   BX487006;
XX
SV   BX487006.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp686L01254_r1  (from clone DKFZp686L01254)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-493
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA   Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by BMFZ (Biomedical Research Center at the
CC   Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
CC   sequencing consortium of the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp686L01254) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .493
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp686L01254"
FT                   /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="adult"
FT                   /tissue_type="cDNA-collection"
XX
SQ   Sequence 493 BP; 125 A; 110 C; 62 G; 196 T; 0 other;

  Query Match             5.3%;  Score 52.6;  DB 2;  Length 493;
  Best Local Similarity   62.0%;  Pred. No. 3e+02;
  Matches  101;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy        820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
```

```
Db        174 GGGAGATGTAGGTCGGAGGATACAAAGTAGCAGATATGTAGGATGAACAAGTC---TAGA 118

Qy        880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939

Db        117 GAAATAATGTACATGAGGACCTCTAGGTAACACAATTGTATTGTATGTGGAATTCATGCT 58

Qy        940 AAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAA 982

Db         57 AAAGGAGTACAATTTAGCTGCTCTTGCAACAAAAACAAAGAAA 15


RESULT 40
BI063151/c
LOCUS       BI063151                606 bp    mRNA    linear   EST 15-JUN-2001
DEFINITION  IL3-UT0117-300301-512-B06 UT0117 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI063151
VERSION     BI063151.1  GI:14470678
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 606)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0117-
            300301-512-B06&t3=2001-03-30&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .606
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UT0117"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      124 a    167 c    143 g    172 t
ORIGIN

  Query Match           5.2%;  Score 51.2;  DB 12;  Length 606;
  Best Local Similarity   62.8%;  Pred. No. 4.3e+02;
  Matches  113;  Conservative    0;  Mismatches   63;  Indels    4;  Gaps    2;

Qy        813 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 872

Db        540 AGGAAACGGGAGATGTAGGTGAAAGGACGCAGAGAGGCAGATATGCAAGATGCATCGCCC 481

Qy        873 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 932

Db        480 AGAG-GCCTAATGTGCATGAGGACTCCTAGTCAACAACATTGA---ATTATACCGGGGAT 425

Qy        933 TCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992

Db        424 TTCTGCTAAGACAGTAGATTTTAGGAGCTCTTGCCATATTGAAAAAAAAAAAAAGGTAAA 365


RESULT 41
BY560588
LOCUS       BY560588                443 bp    mRNA    linear   EST 15-DEC-2002
DEFINITION  BY560588 RIKEN full-length enriched, activated spleen Mus musculus
            cDNA clone F830028K04 3', mRNA sequence.
ACCESSION   BY560588
VERSION     BY560588.1  GI:26895770
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
```

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NOD"
                     /db_xref="taxon:10090"
                     /clone="F830028K04"
                     /tissue_type="activated spleen"
                     /clone_lib="RIKEN full-length enriched, activated spleen"
BASE COUNT       97 a    128 c    117 g    101 t
ORIGIN
```

Query Match 5.1%; Score 50.8; DB 13; Length 443;
Best Local Similarity 62.7%; Pred. No. 5.6e+02;
Matches 79; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        597 CCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCA 656
              || |||||||||||| ||| || |  |||||  |  |  || |||  |||| | |||| |
Db          1 CCAGCCAACTTCTCCCTCCCGCTGTCCCAGACCACTGGTTGTTTCCAGTGCGAAGCTGAA 60

Qy        657 AACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAG 716
              ||| | | |  |||  |  ||| |||| ||    | ||||||  ||  |         |||
Db         61 AACGATGTCGGTGTGGACAGCAGTGCCCGCATCCCGCTGCCCCGAGAGGCCCGAGCCAAG 120

Qy        717 ATGGAG 722
              ||| |
Db        121 CTGGTG 126
```

```
RESULT 42
CNS0091P
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR19D16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN
```

Query Match 4.9%; Score 49; DB 29; Length 925;
Best Local Similarity 15.1%; Pred. No. 7.2e+02;
Matches 47; Conservative 138; Mismatches 127; Indels 0; Gaps 0;

```
Qy        292 CCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCA 351
              |:| :: :|::::| ||: |  |: :  |  : : :    :|::   || :  :   |:
Db        572 CSCCSSCSCSSSSSCSCCBCCCCCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSC 631

Qy        352 CCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGA 411
              ::  :  : ::: ::::|::  : :: : :: :::: : |::     |:   :     |
Db        632 SSSSSCSSSSSSTSSSSTSSSTSKSSSSGSSSSSSSSSYTTSKSTSASGSGSWSAGGGSGS 691

Qy        412 TGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGG 471
              ||: : ::: ::|:|:: ::  |:  :|: : : :: |:::::  ::::     :: ::
Db        692 TGSTSSSSSSSSTSTSSSSVSSGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSS 751

Qy        472 ACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTA 531
               : :   : |   |::: :::|:: : |:: :|: : :::  : ::::   ||: ||::
Db        752 SYSSSTCSCCTCCCSYSYSSSTSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYM 811

Qy        532 TCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACA 591
              . |:  :::|:|:  |||:  :: :||:   |     : :  |: | :  |  ::: |
Db        812 CTCCSTYBMBCYTSTSCGGSSSSSGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCS 871

Qy        592 GGCAGCCTGCCA 603
              ::::::: :::|
Db        872 SSSVSSSSKSSA 883
```

```
RESULT 43
CNS0052P
LOCUS       CNS0052P                 844 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR11P16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      261 a    112 c     92 g     35 t    344 others
ORIGIN
```

```
  Query Match             4.8%;  Score 48;  DB 29;  Length 844;
  Best Local Similarity   17.7%;  Pred. No. 1e+03;
  Matches   75;  Conservative  173;  Mismatches  167;  Indels    8;  Gaps    1;

Qy        534 ACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGG 593
              | :| : : | ::  :: || ||  :|||   :|||  :: ::|::::::   ::   ::
Db        359 AASASVSSACSSVAGSSMAAAGASCBGCAAANMCACGCSSSSSASRSRMMSSCSSCASSS 418

Qy        594 CAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCT 653
              | : : :|:|::  | |:  :|   ::  |: |::| ::: : |:  : ||  : :|:|
Db        419 CSVGSSSCSAMSSSCCCVVSSCGMASSSCCSGGMSASSSSSSGGVSSVSGGRAVGRGRCM 478

Qy        654 GCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAG 713
              | ::|  : :::         | :: |:::: ||: : |: ::: :  :: |: | :::|
Db        479 CCSMVCCCMCSSM--------CCSMCCSVSSVCAVCSSGSVRSVAVGCSVGGRVGGSSRG 530

Qy        714 AAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGG 773
              :||: :|:|| : ::: ::: ::::  |   :: ::: |:: ::   :::|: :  | :
Db        531 RAGRSSGRGGGSSVSSGVVSSSSSVGMGCACSSASSVSCSBSSVASSVSSGBVSRGCGRC 590

Qy        774 AGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTC 833
              :|  :  |: : : : |  :  : :| :   :::::|    :::  :|::  |||::|
Db        591 VGGGVGGGSRVSSCSSGSSSSSSSGCSGVSRSRRSGSANGVRVGGSGRRGGGGGGRRGGA 650

Qy        834 AGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAG 893
              : ::    :|: |:  :  :    | :  :: :::    |::|  :: |:::|::
Db        651 NARADDAAABRABKARANSNADAWGAAVSSSVSSSBGARNGRRAAAVVNGSVSGRBNNKY 710

Qy        894 GACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTT 953
              :  |: :::   |  ::|: :: : :: ::  ::::   : |   |:||  |:::| :
Db        711 KTKTKTRKBKTBTKKKBGKABWKKTKKKKWTRKYKKTWRKGARAAADAAAAGGBKKGGKG 770

Qy        954 TAG 956
              ::|
Db        771 KDG 773
```

```
RESULT 44
AW795001
LOCUS       AW795001                 471 bp    mRNA    linear   EST 16-MAY-2000
DEFINITION  RC6-UM0015-280300-013-C03 UM0015 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW795001
VERSION     AW795001.1  GI:7846871
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC6-UM0015-280
            300-013-C03&t3=2000-03-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="UM0015"
                     /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      180 a     84 c     90 g    117 t
ORIGIN
```

```
  Query Match             4.8%;  Score 47.4;  DB 9;  Length 471;
  Best Local Similarity   59.4%;  Pred. No. 1.6e+03;
  Matches   98;  Conservative    0;  Mismatches   66;  Indels    1;  Gaps    1;

Qy        827 GGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCAC 886
              | | || |||||| | |||| ||||| ||  |||||||||                |
Db         35 GAATGTAAGAGGATGTAAAGTAGCAGACACATAGAATGAAAAAGTGTAGATTTCTTATGT 94

Qy        887 GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAG 946
              | | ||  |    || |  |  | || | |||| ||| ||||||  |||||||
Db         95 ACAACATGAGGACTACAGATAATAATATAGTGCT-GTATCTGGGATTCATGATAAATGAG 153

Qy        947 TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              |   ||||||  | ||  |||||||||  |||||  |||||  ||||
Db        154 TAGATTTTAGTCGGTCCAGCCACAAAAGCAAAAATAAAAACAAAA 198
```

```
RESULT 45
BX339244/c
LOCUS       BX339244                 974 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX339244 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI070YM03 3-PRIME, mRNA sequence.
ACCESSION   BX339244
VERSION     BX339244.1  GI:30335780
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 974)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
```

```
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI070AG02NP1.
FEATURES             Location/Qualifiers
     source          1. .974
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI070YM03"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT       69 a    235 c     80 g    310 t    280 others
ORIGIN

  Query Match             4.8%;  Score 47.2;  DB 13;  Length 974;
  Best Local Similarity   25.8%;  Pred. No. 1.3e+03;
  Matches   85;  Conservative   85;  Mismatches  160;  Indels    0;  Gaps    0;

Qy        663 GCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAG 722
              |:::: | :  ::: :| |:::   ::| ||    |:   :   : |:::| | |:   |
Db        909 GVVMMMGGMGGVMMGVGGGMMMGGGVRGMGGGMGCMGGGSMGMVMGVVVAMAGGMVAMGG 850

Qy        723 GACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGC 782
              |:  ||:|| |   :  | | :| :  : ::  :::   ::|:   :  ||:||: : |:
Db        849 GMRNGGMAGMGMGGMGRGMACMGGMGGRGMMGMVMMVGMSMGSMGGVGGGGMGCMGMGGM 790

Qy        783 CGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGC 842
              :    : |  :||:||:|    : |||    ||: :|||:   |||  |  :| ||::
Db        789 MAGGGVGGGMMAGMGGMGVMVMVMGGGGMKYGGVMMGGGMGVMGGGGMGGGGMGGGGMMNG 730

Qy        843 AAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGG 902
            :|::  :  :  : :::   ::|   :::| :::::   |: : |   : ::|:
Db        729 GGMGMVMGMAAMGGVMGVMMMMNGMMGGGGMSMGGSVMMMMNGGGRGMMAMMCMVVMAVM 670

Qy        903 CCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTC 962
              ::  |: |   ::| |  :  :     |      : |
Db        669 VMCCCMMCMMMMMCCGGMMAMMNMNNNAMMNNNNNAMNNANANMNMANCANAGNNGNNNA 610

Qy        963 TTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
                | ||||||||||||||||||||||||
Db        609 ANAAAAAAAAAAAAAAAAAAAAAAAAAAAA 580


RESULT 46
BZ772737/c
LOCUS       BZ772737                748 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  mcv51f02.g1 HFOSMID007 Homo sapiens genomic, genomic survey
            sequence.
ACCESSION   BZ772737
VERSION     BZ772737.1  GI:28946421
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 748)
  AUTHORS   Cook,L., Delehaunty,K., Fewell,G., Fulton,L., Magrini,V., Mardis,E.
            , Miner,T., Nash,W., Williams,D. and Wilson,R.K.
  TITLE     Homo sapiens Fosmid End Reads
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: mcv51  row: f  column: 02
            Class: fosmid ends
            High quality sequence start: 36
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1. .748
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="HFOSMID007"
                     /note="Vector: pcc01fos; Site_1: Eco72i; Human whole
                     genome fosmid library was prepared at Washington
                     University Genome Sequencing Center. DNA was sheared for
                     blunt-ended ligation into pcc01fos inducible vector. DNA
                     was ordered from Coriell Cell Repository's DNA
                     polymorphism discovery resource."
BASE COUNT      211 a    133 c    104 g    299 t      1 others
ORIGIN

  Query Match             4.7%;  Score 46.4;  DB 29;  Length 748;
  Best Local Similarity   58.1%;  Pred. No. 1.8e+03;
  Matches  100;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        813 AGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCC 872
              |||| | ||   |  |  |  ||||  ||||| |||||  ||||||||| ||| | |
Db        393 AGGAAATGGGGAGATGTAGGTCGAGGGTACAAAGTAGCAGATATGTAGAATAAACAGGTC 334

Qy        873 AGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGT 932
                  |      | | ||  | ||   | ||     |       ||||||| | |
Db        333 TAGAAATGTAATGTATAACATGA-GGGAGATAAACAATAAACTTGCACTGTATTTGCAAT 275

Qy        933 TCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAA 984
              |||| | ||||||||   ||||| ||||| |||||||   |||||||  ||
Db        274 TCATACCAAATGAGTAGATTTTAACTGCTCCTGCCACATGCAAAAAAAGTAA 223


RESULT 47
BX356664/c
LOCUS       BX356664                1201 bp    mRNA    linear   EST 05-MAY-2003
DEFINITION  BX356664 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI015YB03 3-PRIME, mRNA sequence.
ACCESSION   BX356664
VERSION     BX356664.1  GI:30378083
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI015CA02NP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI015YB03"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      116 a     88 c     93 g    398 t    506 others
```

ORIGIN

Query Match 4.7%; Score 46.2; DB 13; Length 1201;
Best Local Similarity 13.1%; Pred. No. 1.6e+03;
Matches 69; Conservative 218; Mismatches 237; Indels 1; Gaps 1;

```
Qy        468 CAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCA 527
Db       1080 SVSSSSVAAVVSAVAAAAVVSSSSSSSSSSSSVSSSSSSSSSMNNSSSSAVASSSSSSSS 1021

Qy        528 CCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGC 587
Db       1020 SSANSSVSSSVASSSAAAASSVAAAVSSAASVSSABAAAAASVAASSSSSSSSSVSSSSM 961

Qy        588 CACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGC 647
Db        960 SSAVSSSSSVSSSSSAVSASSSSSSAAASVSSSAAVSSSVAAASSSSSSSSSVAAASSSS 901

Qy        648 CAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGT 707
Db        900 AAAAVASAVSAAAAVVSWSSVAASSASVASASVSSRSSAANAASSAGSAVAAWASSASSS 841

Qy        708 GACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTC 767
Db        840 SVAVVSAA-AVSSMASAAASSAAVSSSSSSSSAGAVSSSAKSVASSAASVSSAGSGSSAA 782

Qy        768 TACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGG 827
Db        781 VSSVSVRSSSVARVAASSVAAAGSAVASAASSASVASSSVSMAAAAGSAAVSVSASSSAV 722

Qy        828 GAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACG 887
Db        721 SAAAAASSSVSAASAASSSAASSSSASSSASSSAAAVVSSSSSSASSSVARSVAVSSASS 662

Qy        888 GCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGT 947
Db        661 SASSSSAAVAVAAANSANNNSNSSSSSNNNNNAASSSSSANANCNNSSNNNNNSSSSSNN 602

Qy        948 GTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
Db        601 NSSNSNANNNNANAAAVNAAANAAAANAAASAAAAAAAAAAAAA 557
```

RESULT 48
CB615918/c
LOCUS CB615918 569 bp mRNA linear EST 07-APR-2003
DEFINITION AMGNNUC:URRG1-00010-E11-A urrg1 (14046) Rattus norvegicus cDNA clone urrg1-00010-e11 5', mRNA sequence.
ACCESSION CB615918
VERSION CB615918.1 GI:29575806
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 569)
AUTHORS Amgen EST Program.
TITLE Amgen Rat EST Program
JOURNAL Unpublished
COMMENT Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00010 row: e column: 11.
FEATURES Location/Qualifiers
source 1. .569
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="urrg1-00010-e11"
/clone_lib="urrg1 (14046)"
/note="Vector: pT7T3D-PAC"

BASE COUNT 136 a 133 c 128 g 172 t
ORIGIN

Query Match 4.6%; Score 45.6; DB 14; Length 569;
Best Local Similarity 55.8%; Pred. No. 2.7e+03;
Matches 87; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy        836 AGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGA 895
Db        178 AGAACGGACGGCCGCTGTGTAGCTCCAGTCACCATACAAAGATGCCAGTTCTACAGAGTG 119

Qy        896 CTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTA 955
Db        118 GAAGTGCCTAGCTTTGAGCTGTGTATATAACTTAAGAATGTTCAAACTAAGATTATATTA 59

Qy        956 GCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 991
Db         58 AAAACACATGACATAAAAAAAAAAAAAAAAAAAAAAA 23
```

RESULT 49
CD397038/c
LOCUS CD397038 626 bp mRNA linear EST 01-JUN-2003
DEFINITION Gm_ck17710 Soybean induced by Salicylic Acid Glycine max cDNA 3', mRNA sequence.
ACCESSION CD397038
VERSION CD397038.1 GI:31311835
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 626)
AUTHORS Tian,A.-G., Wang,J., Cui,P., Han,Y.-J., Xu,H., Cong,L.-J., Huang,X.-G., Wang,X.-L., Jiao,Y.-Z., Wang,B.-J., Wang,Y.-J., Zhang,J.-S., Chen,S.-Y. and Yu,J.
TITLE Soybean Expressed Sequence Tags Sequencing
JOURNAL Unpublished
COMMENT Contact: Chen S-Y
Plant Biotechnology Laboratory
Institute of Genetics and Developmental Biology, CAS, China
Datun road, Beijing 100101, China
Tel: 86-10-64886859
Fax: 86-10-64873428
Email: sychen@genetics.ac.cn
Email: sychen@genetics.ac.cn
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .626
/organism="Glycine max"
/mol_type="mRNA"
/cultivar="Kefeng 1"
/db_xref="taxon:3847"
/tissue_type="Seedlings"
/dev_stage="two-week seedlings"
/lab_host="XL1-Blue MRF' strain"
/clone_lib="Soybean induced by Salicylic Acid"
/note="Vector: pBluescript SK+; Site_1: EcoR I; Site_2: Xho I; The cDNA library was constructed by He, C-Y from mRNA isolated from two-week seedlings (cultivar Kefeng 1) treated by spraying 2.0mM salicylic acid for 24, 36, 48 and 72 h. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI striction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL1-Blue MRF' host cells (Stratagene)."

BASE COUNT 164 a 152 c 145 g 165 t
ORIGIN

```
  Query Match             4.6%;  Score 45.6;  DB 14;  Length 626;
  Best Local Similarity   71.4%;  Pred. No. 2.6e+03;
  Matches   60;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy        908 AGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCC 967
              || || |||| || | | ||  ||| |   |||| |||| || |  | |||| |
Db         91 AGAGTCCACTCATCTTCTGTGTTGTTAAAAAAAAACCCGTGTGGTTAATTTCCTCTTCCA 32

Qy        968 ACAAAAAAAAAAAAAAAAAAAAAA 991
              | ||||||||||||||||||||||
Db         31 AAAAAAAAAAAAAAAAAAAAAAAA 8


RESULT 50
CNS006XK
LOCUS       CNS006XK                 935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR14N09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
ORIGIN

  Query Match             4.6%;  Score 45.6;  DB 29;  Length 935;
  Best Local Similarity   25.8%;  Pred. No. 2.1e+03;
  Matches  109;  Conservative  105;  Mismatches  208;  Indels    0;  Gaps    0;

Qy        342 GACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCC 401
              ::| |:: :  |  |  : ::::: |: : ||| :| :  : ::|: |   :   | ||:
Db        513 SMCRTSSGSTGCSTCCMYMSSSVSCCSCSGTCCGYCSSCTSCKMCSCTYGCKCGCGCGCS 572

Qy        402 AGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTC 461
               : :  :   ::: :| :  :||: :: |:  : | |      | || |||:||   |
Db        573 TSCSSSSCCSBBSYSTCSCTBCTKCSSGCSTGSGCTGCCGGGGSGCGGCGSGCGCGCGSG 632

Qy        462 ACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGC 521
```

```
              || ::   : :::|  :|::|  ::: :| |  |  | :  ||||  : :| |:
Db        633 GGCGGGCSSSGGSSGSSSGGSSGSSCGSSSSGSGGGGSGGGCGSGGCGGCGCGSCSGCGS 692

Qy        522 AGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGA 581
              ||:: |        |    ||     : ::   |   ::  :| :| :| |: :| : :
Db        693 CGCSSGCGSGCGSGCGSGGGCGSCGCSCSSCGCGSGCSSGSSGCSCCSCCGSCSCSSSSG 752

Qy        582 CCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTC 641
              :|    | |  :: : | |::     : ||  :: |||: ||| | |  ||  : :   |
Db        753 SCCSCGCGCGMSMCSGCGGSSCGGCCSGCCGCSSCGCCSGGCCSGCCGGCGCGGSGSCGGC 812

Qy        642 TGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCA 701
              | ::: |:   |      ||:|   | |  :: | ||||  :   :   :|   : |
Db        813 GCGGSSSGGSGSCCGSCGGGCGSCCSGGGCGCSSSCSGCGCGSGGSSCCGSGGCGSGCGS 872

Qy        702 GGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTG 761
              |: || : :: :  :   | |:   : : ||| ||:|  | : ||::|  :|   ||  |
Db        873 GSSGGGSSSSGSGSSGGCGSGSGGGSCSGGGGSCCSCSCGCSCGCSSCSSSCSCCCCCCG 932

Qy        762 CC 763
              :|
Db        933 SC 934


Search completed: February  9, 2004, 14:58:47
Job time : 2542 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 13:04:22 ; Search time 164 Seconds
(without alignments)
2669.830 Million cell updates/sec

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggcacgagccaggaactagg.....aaaaaaaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	992	100.0	992	4	US-09-996-243-222	Sequence 222, App
c	2	47	4.7	87350	3	US-08-781-891-79	Sequence 79, Appl
c	3	47	4.7	87350	4	US-09-618-166-79	Sequence 79, Appl
c	4	47	4.7	87543	4	US-09-791-211-3	Sequence 3, Appli
	5	43.6	4.4	289	3	US-09-007-005-17	Sequence 17, Appl
	6	43.6	4.4	289	3	US-09-244-796-17	Sequence 17, Appl
c	7	41.4	4.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c	8	40.6	4.1	480	4	US-09-252-991A-11640	Sequence 11640, A
	9	40.6	4.1	1575	4	US-09-252-991A-11799	Sequence 11799, A
	10	40.6	4.1	2022	4	US-09-252-991A-11716	Sequence 11716, A
	11	40.2	4.1	1332	4	US-09-252-991A-10993	Sequence 10993, A
c	12	40.2	4.1	1656	4	US-09-252-991A-11258	Sequence 11258, A
c	13	40.2	4.1	2217	4	US-09-252-991A-11190	Sequence 11190, A
c	14	37.8	3.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	15	37.8	3.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	16	37.6	3.8	1048	4	US-09-522-714-21	Sequence 21, Appl
	17	37.6	3.8	1396	1	US-08-123-161A-11	Sequence 11, Appl
	18	37.6	3.8	1396	1	US-08-483-278-11	Sequence 11, Appl
c	19	37.6	3.8	1992	4	US-09-252-991A-5567	Sequence 5567, Ap
c	20	37.4	3.8	520	2	US-08-997-080-179	Sequence 179, App
c	21	37.4	3.8	520	2	US-08-997-362-179	Sequence 179, App
c	22	37.4	3.8	520	3	US-09-095-855-179	Sequence 179, App
c	23	37.4	3.8	520	4	US-09-324-542-179	Sequence 179, App
c	24	37.4	3.8	520	4	US-09-205-426-179	Sequence 179, App
c	25	37.4	3.8	570	3	US-09-095-855-202	Sequence 202, App
c	26	37.4	3.8	570	4	US-09-205-426-202	Sequence 202, App
	27	36.6	3.7	1273	3	US-09-319-892-3	Sequence 3, Appli
	28	36.6	3.7	1273	4	US-09-727-238-3	Sequence 3, Appli
	29	36.4	3.7	801	4	US-09-252-991A-11102	Sequence 11102, A
	30	36.4	3.7	1116	4	US-09-252-991A-11158	Sequence 11158, A
c	31	36.4	3.7	1584	4	US-09-252-991A-11094	Sequence 11094, A
	32	36.4	3.7	2028	2	US-09-211-930-12	Sequence 12, Appl
	33	36.4	3.7	2028	3	US-09-340-993-12	Sequence 12, Appl
	34	36.4	3.7	2028	4	US-09-468-442-12	Sequence 12, Appl
c	35	36.2	3.6	570	4	US-09-252-991A-5589	Sequence 5589, Ap
c	36	36.2	3.6	1545	4	US-09-252-991A-5553	Sequence 5553, Ap
	37	36.2	3.6	1914	4	US-09-252-991A-5670	Sequence 5670, Ap
	38	36	3.6	2625	3	US-09-245-041-18	Sequence 18, Appl
	39	36	3.6	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	40	36	3.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	41	35.8	3.6	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	42	35.8	3.6	28958	1	US-08-456-837-6	Sequence 6, Appli
c	43	35.8	3.6	28958	1	US-08-457-342-6	Sequence 6, Appli
c	44	35.8	3.6	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	45	35.8	3.6	28958	1	US-08-458-076A-6	Sequence 6, Appli
c	46	35.8	3.6	28958	1	US-08-764-233A-4	Sequence 4, Appli
c	47	35.8	3.6	28958	1	US-08-457-335A-6	Sequence 6, Appli
c	48	35.8	3.6	28958	1	US-08-729-214-6	Sequence 6, Appli
c	49	35.8	3.6	28958	3	US-09-028-934-6	Sequence 6, Appli
c	50	35.8	3.6	49377	1	US-08-764-233A-1	Sequence 1, Appli
	51	35.6	3.6	366	4	US-09-775-932-13	Sequence 13, Appl
	52	35.6	3.6	387	4	US-09-775-932-11	Sequence 11, Appl
	53	35.6	3.6	581	2	US-08-461-030C-1	Sequence 1, Appli
	54	35.6	3.6	588	3	US-08-744-138-1	Sequence 1, Appli
	55	35.6	3.6	588	4	US-09-241-376-1	Sequence 1, Appli
c	56	35.4	3.6	294	4	US-09-252-991A-10962	Sequence 10962, A
	57	35.4	3.6	447	4	US-09-252-991A-8928	Sequence 8928, Ap
c	58	35.4	3.6	676	2	US-08-533-669A-3	Sequence 3, Appli
c	59	35.4	3.6	676	4	US-09-183-861-3	Sequence 3, Appli
c	60	35.4	3.6	676	4	US-09-022-765-3	Sequence 3, Appli
c	61	35.4	3.6	676	4	US-09-551-974A-3	Sequence 3, Appli
c	62	35.4	3.6	696	4	US-09-252-991A-9236	Sequence 9236, Ap
	63	35.4	3.6	1011	4	US-09-252-991A-8814	Sequence 8814, Ap
c	64	35.4	3.6	1125	4	US-09-252-991A-14001	Sequence 14001, A
	65	35.4	3.6	1617	4	US-09-252-991A-14135	Sequence 14135, A
c	66	35.4	3.6	1728	4	US-09-252-991A-9152	Sequence 9152, Ap
c	67	35.2	3.5	23673	4	US-09-773-816-1	Sequence 1, Appli
	68	35	3.5	588	4	US-09-205-258-64	Sequence 64, Appl
	69	35	3.5	1683	3	US-09-347-803-11	Sequence 11, Appl
	70	34.8	3.5	1646	4	US-09-443-041A-19	Sequence 19, Appl
	71	34.8	3.5	1969	1	US-07-937-609-28	Sequence 28, Appl
	72	34.8	3.5	1969	3	US-08-029-170-28	Sequence 28, Appl
c	73	34.6	3.5	816	4	US-09-252-991A-11434	Sequence 11434, A
	74	34.6	3.5	1116	4	US-09-522-714-31	Sequence 31, Appl
	75	34.6	3.5	1230	4	US-09-461-325-80	Sequence 80, Appl
	76	34.6	3.5	1254	4	US-09-252-991A-11478	Sequence 11478, A
	77	34.6	3.5	1524	4	US-09-252-991A-11455	Sequence 11455, A
c	78	34.6	3.5	2028	4	US-09-252-991A-11416	Sequence 11416, A
	79	34.6	3.5	2346	3	US-09-193-503B-5	Sequence 5, Appli
	80	34.6	3.5	2346	4	US-09-415-839-5	Sequence 5, Appli
	81	34.6	3.5	3471	5	PCT-US93-00227-2	Sequence 2, Appli
	82	34.6	3.5	3530	3	US-08-704-711A-10	Sequence 10, Appl
	83	34.6	3.5	3530	4	US-09-521-220-10	Sequence 10, Appl
	84	34.6	3.5	4190	3	US-08-938-291A-2	Sequence 2, Appli
	85	34.6	3.5	4190	4	US-09-589-619-2	Sequence 2, Appli
	86	34.2	3.4	423	1	US-08-470-179-186	Sequence 186, App
c	87	34.2	3.4	1863	4	US-09-252-991A-3225	Sequence 3225, Ap
	88	34.2	3.4	2154	4	US-09-252-991A-3346	Sequence 3346, Ap
	89	34.2	3.4	9208	4	US-09-068-506-1	Sequence 1, Appli
	90	34	3.4	2461	1	US-08-282-141-1	Sequence 1, Appli
c	91	34	3.4	3402	4	US-09-252-991A-1374	Sequence 1374, Ap
	92	34	3.4	3687	4	US-09-252-991A-1193	Sequence 1193, Ap
	93	34	3.4	4266	4	US-09-252-991A-1234	Sequence 1234, Ap
c	94	33.8	3.4	1518	3	US-09-240-915-2	Sequence 2, Appli
c	95	33.8	3.4	1518	3	US-09-591-435-2	Sequence 2, Appli
c	96	33.8	3.4	1536	4	US-09-252-991A-11652	Sequence 11652, A
	97	33.8	3.4	1656	4	US-09-228-986-63	Sequence 63, Appl
	98	33.8	3.4	1761	4	US-09-252-991A-11704	Sequence 11704, A
	99	33.8	3.4	1767	4	US-09-252-991A-7614	Sequence 7614, Ap
	100	33.8	3.4	2118	4	US-09-252-991A-11782	Sequence 11782, A

c	101	33.8	3.4	3447	2	US-08-313-185-57 Sequence 57, Appl
c	102	33.8	3.4	3447	3	US-09-082-614A-57 Sequence 57, Appl
	103	33.6	3.4	2572	3	US-08-804-439A-9 Sequence 9, Appli
	104	33.6	3.4	2572	3	US-08-720-229-9 Sequence 9, Appli
	105	33.6	3.4	2967	4	US-09-185-501B-12 Sequence 12, Appl
c	106	33.6	3.4	15652	4	US-09-422-936-60 Sequence 60, Appl
	107	33.4	3.4	407	4	US-08-948-113D-11 Sequence 11, Appl
	108	33.4	3.4	1053	4	US-09-252-991A-14223 Sequence 14223, A
c	109	33.4	3.4	1846	1	US-08-483-389-117 Sequence 117, App
c	110	33.4	3.4	10660	2	US-08-267-803B-8 Sequence 8, Appli
c	111	33.4	3.4	10660	3	US-09-041-886-16 Sequence 16, Appl
c	112	33.2	3.3	465	4	US-09-252-991A-12679 Sequence 12679, A
	113	33.2	3.3	750	4	US-09-252-991A-13279 Sequence 13279, A
	114	33.2	3.3	858	4	US-09-252-991A-8589 Sequence 8589, Ap
c	115	33.2	3.3	996	4	US-09-252-991A-12822 Sequence 12822, A
	116	33.2	3.3	1897	1	US-08-184-632-1 Sequence 1, Appli
	117	33.2	3.3	1947	1	US-07-951-715A-3 Sequence 3, Appli
	118	33.2	3.3	1947	2	US-08-459-448A-3 Sequence 3, Appli
	119	33.2	3.3	1947	3	US-08-459-595A-3 Sequence 3, Appli
	120	33.2	3.3	1947	3	US-08-459-504B-3 Sequence 3, Appli
	121	33.2	3.3	1947	3	US-08-459-444-3 Sequence 3, Appli
	122	33.2	3.3	1947	4	US-09-547-422-3 Sequence 3, Appli
	123	33.2	3.3	2190	2	US-08-492-027A-7 Sequence 7, Appli
	124	33.2	3.3	2194	2	US-08-948-569A-9 Sequence 9, Appli
	125	33.2	3.3	2194	2	US-09-188-469-9 Sequence 9, Appli
	126	33.2	3.3	2194	3	US-09-397-238A-9 Sequence 9, Appli
c	127	33.2	3.3	2502	4	US-09-252-991A-8656 Sequence 8656, Ap
	128	33.2	3.3	2852	3	US-09-027-137-2 Sequence 2, Appli
	129	33.2	3.3	2852	4	US-09-344-441-2 Sequence 2, Appli
	130	33.2	3.3	3468	1	US-07-951-715A-2 Sequence 2, Appli
	131	33.2	3.3	3468	1	US-07-951-715A-4 Sequence 4, Appli
	132	33.2	3.3	3468	1	US-07-951-715A-8 Sequence 8, Appli
	133	33.2	3.3	3468	2	US-08-459-448A-2 Sequence 2, Appli
	134	33.2	3.3	3468	2	US-08-459-448A-4 Sequence 4, Appli
	135	33.2	3.3	3468	2	US-08-459-448A-8 Sequence 8, Appli
	136	33.2	3.3	3468	3	US-08-459-595A-2 Sequence 2, Appli
	137	33.2	3.3	3468	3	US-08-459-595A-4 Sequence 4, Appli
	138	33.2	3.3	3468	3	US-08-459-595A-8 Sequence 8, Appli
	139	33.2	3.3	3468	3	US-08-459-504B-2 Sequence 2, Appli
	140	33.2	3.3	3468	3	US-08-459-504B-4 Sequence 4, Appli
	141	33.2	3.3	3468	3	US-08-459-504B-8 Sequence 8, Appli
	142	33.2	3.3	3468	3	US-08-459-444-2 Sequence 2, Appli
	143	33.2	3.3	3468	3	US-08-459-444-4 Sequence 4, Appli
	144	33.2	3.3	3468	3	US-08-459-444-8 Sequence 8, Appli
	145	33.2	3.3	3468	3	US-09-053-549-3 Sequence 3, Appli
	146	33.2	3.3	3468	3	US-09-053-549-5 Sequence 5, Appli
	147	33.2	3.3	3468	4	US-09-547-422-2 Sequence 2, Appli
	148	33.2	3.3	3468	4	US-09-547-422-4 Sequence 4, Appli
	149	33.2	3.3	3468	4	US-09-547-422-8 Sequence 8, Appli
	150	33.2	3.3	3546	1	US-07-951-715A-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-996-243-222
; Sequence 222, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi,Avi J.
; APPLICANT: Baker,Kevin P.
; APPLICANT: Botstein,David
; APPLICANT: Desnoyers,Luc
; APPLICANT: Eaton,Dan L.
; APPLICANT: Ferrara,Napoleone
; APPLICANT: Fong,Sherman
; APPLICANT: Gerber,Hanspeter
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,J.Christopher
; APPLICANT: Gurney,Austin L.
; APPLICANT: Kljavin,Ivar J.
; APPLICANT: Napier,Mary A.
; APPLICANT: Pan,James
; APPLICANT: Paoni,Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C13
; CURRENT APPLICATION NUMBER: US/09/996,243
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738

; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 100.0%; Score 992; DB 4; Length 992;
Best Local Similarity 100.0%; Pred. No. 1.4e-243;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 2
US-08-781-891-79/c
; Sequence 79, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  79:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  87350 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-79

  Query Match             4.7%;  Score 47;  DB 3;  Length 87350;
  Best Local Similarity   78.9%;  Pred. No. 0.045;
  Matches   56;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
              |||||||| ||| ||| ||||||||| ||| ||||||| ||||||||||| ||||||
Db      18894 GTATTTGGGATTCCTGCTAAATGAGTGGATTTCAGCTGCTTTTGCCACAAAAACAAAAAA 18835

Qy        982 AAAAAAAAAAA 992
              |    || ||
Db      18834 ATGGTTAACAA 18824


RESULT 3
US-09-618-166-79/c
; Sequence 79, Application US/09618166
; Patent No. 6583112
;  GENERAL INFORMATION:
;       APPLICANT: Fu, Ying-Hui
;                  Yu, Chang-En
;                  Oshima, Junko
;                  Mulligan, John T.
;                  Schellenberg, Gerald D.
;       TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;                           WERNER'S SYNDROME
;       NUMBER OF SEQUENCES: 209
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed Intellectual Property Law Group
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/618,166
;            FILING DATE: 17-Jul-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: McMasters, David D.
;            REGISTRATION NUMBER: 33,963
;            REFERENCE/DOCKET NUMBER: 240052.419C1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;  INFORMATION FOR SEQ ID NO: 79:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 87350 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-618-166-79

  Query Match             4.7%;  Score 47;  DB 4;  Length 87350;
  Best Local Similarity   78.9%;  Pred. No. 0.045;
  Matches   56;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
```

```
                ||||||||  ||| ||| |||||||||  ||| |||||||| ||||||||||||| ||||||
Db      18894 GTATTTGGGATTCCTGCTAAATGAGTGGATTTCAGCTGCTTTTGCCACAAAAACAAAAAA 18835

Qy        982 AAAAAAAAAA 992
                |     || ||
Db      18834 ATGGTTAACAA 18824


RESULT 4
US-09-791-211-3/c
; Sequence 3, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 3
;   LENGTH: 87543
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 7421
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 7427
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 11609
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 12605
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 12742
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 29370
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 29422
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 29979
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 29980
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 29981
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 30136
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 30140
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 31205
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 31206
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 31592
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 33095
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 33160
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 34066
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 34072
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 36816
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 39020
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 42164
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 42459
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 46808
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 46823
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 46826
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 47291
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 52786
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 52787
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 53384
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 54684
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 59215
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 59235
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 59242
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 63290
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 66614
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 68660
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 68697
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 68718
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 68733
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
```

```
;   LOCATION: 68739
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 69785
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 79134
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 79198
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 86336
;   OTHER INFORMATION: unknown
;   OTHER INFORMATION:
US-09-791-211-3

  Query Match             4.7%;  Score 47;  DB 4;  Length 87543;
  Best Local Similarity   78.9%;  Pred. No. 0.045;
  Matches   56;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAA 981
              |||||||| ||| ||| |||||||||  ||| |||||||| |||||||||||| ||||||
Db      19087 GTATTTGGGATTCCTGCTAAATGAGTGGATTTCAGCTGCTTTTGCCACAAAAACAAAAAA 19028

Qy        982 AAAAAAAAAAA 992
              |     || ||
Db      19027 ATGGTTAACAA 19017


RESULT 5
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             4.4%;  Score 43.6;  DB 3;  Length 289;
  Best Local Similarity   12.1%;  Pred. No. 0.041;
  Matches   27;  Conservative   85;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        769 ACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGG 828
              : ::: : ::  :  :    ::  : : : :    : :::   :::   ::: : :::
Db         65 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 124

Qy        829 AGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGG 888
              :  :   : ::: : : :   :::   : : : :   :  ::   ::: :  :: : :::
Db        125 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 184

Qy        889 CAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTG 948
              : : :  :  :  ::   ::: : ::     :      |: ::  : : : :|  :
Db        185 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARARCRU 244

Qy        949 TGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              : ::|  :  | :   | ||||||||||||||||||||||
Db        245 RCRURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAAAAAAAA 287


RESULT 6
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match             4.4%;  Score 43.6;  DB 3;  Length 289;
  Best Local Similarity   12.1%;  Pred. No. 0.041;
  Matches   27;  Conservative   85;  Mismatches  111;  Indels    0;  Gaps    0;

Qy        769 ACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGG 828
              : ::: : ::  :  :    ::  : : : :    : :::   :::   ::: : :::
Db         65 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 124

Qy        829 AGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGG 888
              :  :   : ::: : : :   :::   : : : :   :  ::   ::: :  :: : :::
Db        125 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRN 184

Qy        889 CAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTG 948
              : : :  :  :  ::   ::: : ::     :      |: ::  : : : :|  :
Db        185 RNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRCRARGRCRURGRCRGRURARARCRU 244

Qy        949 TGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              : ::|  :  | :   | ||||||||||||||||||||||
Db        245 RCRURURGRCRGRCRUAAAAAAAAAAAAAAAAAAAAAAAAAA 287


RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Foley & Lardner
;       STREET:  1800 Diagonal Road, Suite 500
;       CITY:  Alexandria
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22313-0299
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/232,463
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/935,313
;       FILING DATE:
;       APPLICATION NUMBER:  EP 91 114 300.6
;       FILING DATE:  26-AUG-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  7218 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     IMMEDIATE SOURCE:
;       CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             4.2%;  Score 41.4;  DB 1;  Length 7218;
  Best Local Similarity   1.7%;  Pred. No. 0.48;
  Matches    6;  Conservative  203;  Mismatches  144;  Indels    0;  Gaps    0;

Qy        640 TCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCC 699
              | |||| | ::: : ::: :: :  :: :   :: : :: :     : :: :: :
Db       1442 TTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1383

Qy        700 CAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCT 759
              ::: :: ::  :::::: :::::  :: ::::      ::::::    :    :
Db       1382 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1323

Qy        760 TGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAG 819
               :  :    : ::::: :   :  :   ::: :::::::::   ::::::   :::: ::
Db       1322 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1263

Qy        820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
              :::: :::::::  :::::: : :::: :: ::  : : :::: :::  :   ::::::
Db       1262 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1203

Qy        880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939
              ::: : :: ::::::  : :::  :  :: : : :  :   : :   :::::    : : :
Db       1202 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1143

Qy        940 AAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ::: ::: : :    ::  :     :  : ::::::::::::::::::::::
Db       1142 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1090


RESULT 8
US-09-252-991A-11640/c
; Sequence 11640, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11640
;   LENGTH:  480
;   TYPE:  DNA
;   ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-11640

  Query Match             4.1%;  Score 40.6;  DB 4;  Length 480;
  Best Local Similarity   44.4%;  Pred. No. 0.29;
  Matches  163;  Conservative    0;  Mismatches  204;  Indels    0;  Gaps    0;

Qy        260 GAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAA 319
              | | ||||||  | |||   ||  ||||     ||  ||  | |  || | |  |  | |
Db        424 GGAGATCAAGAAGACCATCGAGCAGGTGGTCGCCAGCGGCCTGATCTACCTGCCCTCCGG 365

Qy        320 CGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTC 379
              |||      |  |  || |||  | ||  ||  ||| ||| |     |  |  |  |||
Db        364 CGTTCGCGACCTGCACAATCCGCAACTCGACAAGTACCTCTCCACCTATTGCCGCGGCTC 305

Qy        380 AGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGT 439
              || | |   |  || | |  || |  |||| | |  |  ||| |||  |      |
Db        304 CGGCGGCATGGGCCACCGGGAGCGGATCAAGATCCTCAAGCTGCTCTGGGACGCCATCGG 245

Qy        440 GTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGAT 499
              |||  | | |  |  |  ||| | ||  |          | |  |||   |||
Db        244 CAGCGAGTTCGGCGGCCGCCACGAGCTGTACGAGATCAACTACGCCGGCAGCCAGGACGA 185

Qy        500 GATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGG 559
              |||| ||  |  | ||  | ||     ||  |       |||  ||  | |  ||
Db        184 GATCCGCATGCAGGCCCTGCGCCAGGCGGTCGGCAGCGGGGCGATGAAGGGCATGCTCGG 125

Qy        560 GCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCC 619
              ||||  | | |    | |   | | ||  | |       | | |  |  |||||
Db        124 CATGGTCGAGCAGTGCATGGGCGACTACGACGAGAACGGCTGGACCGTGCCGCACCTGCA 65

Qy        620 GAGCCAG 626
              |  | | |
Db         64 CAACCCG 58


RESULT 9
US-09-252-991A-11799
; Sequence 11799, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11799
;   LENGTH:  1575
;   TYPE:  DNA
;   ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-11799

  Query Match             4.1%;  Score 40.6;  DB 4;  Length 1575;
```

Best Local Similarity 44.4%; Pred. No. 0.44;
Matches 163; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy        260 GAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAA 319
Db       1173 GGAGATCAAGAAGACCATCGAGCAGGTGGTCGCCAGCGGCCTGATCTACCTGCCCTCCGG 1232

Qy        320 CGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTC 379
Db       1233 CGTTCGCGACCTGCACAATCCGCAACTCGACAAGTACCTCTCCACCTATTGCCGCGGCTC 1292

Qy        380 AGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGT 439
Db       1293 CGGCGGCATGGGCCACCGGGAGCGGATCAAGATCCTCAAGCTGCTCTGGGACGCCATCGG 1352

Qy        440 GTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGAT 499
Db       1353 CAGCGAGTTCGGCGGCCGCCACGAGCTGTACGAGATCAACTACGCCGGCAGCCAGGACGA 1412

Qy        500 GATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGG 559
Db       1413 GATCCGCATGCAGGCCCTGCGCCAGGCGGTCGGCAGCGGGGCGATGAAGGGCATGCTCGG 1472

Qy        560 GCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCC 619
Db       1473 CATGGTCGAGCAGTGCATGGGCGACTACGACGAGAACGGCTGGACCGTGCCGCACCTGCA 1532

Qy        620 GAGCCAG 626
Db       1533 CAACCCG 1539
```

RESULT 10
US-09-252-991A-11716
; Sequence 11716, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11716
; LENGTH: 2022
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11716

Query Match 4.1%; Score 40.6; DB 4; Length 2022;
Best Local Similarity 44.4%; Pred. No. 0.48;
Matches 163; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy        260 GAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAA 319
Db       1078 GGAGATCAAGAAGACCATCGAGCAGGTGGTCGCCAGCGGCCTGATCTACCTGCCCTCCGG 1137

Qy        320 CGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTC 379
Db       1138 CGTTCGCGACCTGCACAATCCGCAACTCGACAAGTACCTCTCCACCTATTGCCGCGGCTC 1197

Qy        380 AGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGT 439
Db       1198 CGGCGGCATGGGCCACCGGGAGCGGATCAAGATCCTCAAGCTGCTCTGGGACGCCATCGG 1257

Qy        440 GTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGAT 499
Db       1258 CAGCGAGTTCGGCGGCCGCCACGAGCTGTACGAGATCAACTACGCCGGCAGCCAGGACGA 1317

Qy        500 GATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGG 559
Db       1318 GATCCGCATGCAGGCCCTGCGCCAGGCGGTCGGCAGCGGGGCGATGAAGGGCATGCTCGG 1377

Qy        560 GCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCC 619
Db       1378 CATGGTCGAGCAGTGCATGGGCGACTACGACGAGAACGGCTGGACCGTGCCGCACCTGCA 1437

Qy        620 GAGCCAG 626
Db       1438 CAACCCG 1444
```

RESULT 11
US-09-252-991A-10993
; Sequence 10993, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10993
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10993

Query Match 4.1%; Score 40.2; DB 4; Length 1332;
Best Local Similarity 49.8%; Pred. No. 0.53;
Matches 102; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```
Qy        218 CCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAA 277
Db        826 CCCGCCACCGCAGAGGATCGCCGGGATCCGCGGGATGCCGAATACCACCTCGGTGACCGA 885

Qy        278 GAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAG 337
Db        886 GCCGATGGCGATGGTCGCGGTGACCGCCACCATCAGGTCGTAGGCGACCATGACCACCCC 945

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
Db        946 GCCGATCAGGCGCGAGGAGGCCTGGTAGCGCTGCTCCAGCACCTGGGTCACGGTGAACAC 1005

Qy        398 TGCCAGGCTACAGATGCACTGGGAG 422
Db       1006 CCGCAGGCGCGCGATCTGCCGCGAG 1030
```

RESULT 12
US-09-252-991A-11258/c
; Sequence 11258, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 11258
;   LENGTH: 1656
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11258

  Query Match             4.1%;  Score 40.2;  DB 4;  Length 1656;
  Best Local Similarity   49.8%;  Pred. No. 0.57;
  Matches  102;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        218 CCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAA 277
              || |||||| | |   ||| ||     || | |     | || | || |  |||| || |
Db        570 CCCGCCACCGCAGAGGATCGCCGGGATCCGCGGGATGCCGAATACCACCTCGGTGACCGA 511

Qy        278 GAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAG 337
              | | ||| || |  |   | | | ||| || |||     | |  ||  | |   ||
Db        510 GCCGATGGCGATGGTCGCGGTGACCGCCACCATCAGGTCGTAGGCGACCATGACCACCCC 451

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
              ||   |  ||  |   |||   |  | |||||  | |  | | |   ||| |||
Db        450 GCCGATCAGGCGCGAGGAGGCCTGGTAGCGCTGCTCCAGCACCTGGGTCACGGTGAACAC 391

Qy        398 TGCCAGGCTACAGATGCACTGGGAG 422
              |||||   |||  | | |||
Db        390 CCGCAGGCGCGCGATCTGCCGCGAG 366


RESULT 13
US-09-252-991A-11190/c
; Sequence 11190, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11190
;   LENGTH: 2217
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (2217)
;   OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-11190

  Query Match             4.1%;  Score 40.2;  DB 4;  Length 2217;
  Best Local Similarity   49.8%;  Pred. No. 0.63;
  Matches  102;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        218 CCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAA 277
              || |||||| | |   ||| ||     || | |     | || | || |  |||| || |
Db        778 CCCGCCACCGCAGAGGATCGCCGGGATCCGCGGGATGCCGAATACCACCTCGGTGACCGA 719

Qy        278 GAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAG 337
              | | ||| || |  |   | | | ||| || |||     | |  ||  | |   ||
Db        718 GCCGATGGCGATGGTCGCGGTGACCGCCACCATCAGGTCGTAGGCGACCATGACCACCCC 659

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
              ||   |  ||  |   |||   |  | |||||  | |  | | |   ||| |||
Db        658 GCCGATCAGGCGCGAGGAGGCCTGGTAGCGCTGCTCCAGCACCTGGGTCACGGTGAACAC 599

Qy        398 TGCCAGGCTACAGATGCACTGGGAG 422
              |||||   |||  | | |||
Db        598 CCGCAGGCGCGCGATCTGCCGCGAG 574


RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             3.8%;  Score 37.8;  DB 3;  Length 4403765;
  Best Local Similarity   47.3%;  Pred. No. 34;
  Matches  114;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
              | | |||||  ||  |  | || ||  ||||| | |  || | |  | | ||  || ||
Db     284113 TTCCGACCTTGTCGACATCCTCGGCTCGGCGTACGCGGCCGCCGACGGCGATCCGGCCAC 284054

Qy        398 TGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAA 457
              |||  ||      |  |||  |||  | ||      |||  |   | |  ||||  |   |
Db     284053 CGCCTGGACCGCGCCGCAGCGGGTGGTGCAACACAAAACCCCGCCGACGCTGACGCTGAC 283994

Qy        458 CTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTC 517
              | |  | | ||     |  | | | ||  |  || |||     | | |   ||   |
Db     283993 CCTGCCGCGGCCCACCGTGGTGACCGGGCTGCGGCTGGCCGCCAGCCGGTCGATGTTGCC 283934

Qy        518 GGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCA 577
              || |   ||  | |      || |    | ||  | |      |||||||  |  | |||
Db     283933 GGCCCATCCGACGGTGGTGGCCATCAACCTGGGCGACGGCCCCCAGGTCCGACAACTGCA 283874

Qy        578 G 578
              |
Db     283873 G 283873


RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
```

```
US-09-103-840A-1

  Query Match             3.8%;  Score 37.8;  DB 3;  Length 4411529;
  Best Local Similarity   47.3%;  Pred. No. 34;
  Matches  114;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
Db     284000 TTCCGACCTTGTCGACATCCTCGGCTCGGCGTACGCGGCCGCCGACGGCGATCCGGCCAC 283941
Qy        398 TGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAA 457
Db     283940 CGCCTGGACCGCGCCGCAGCGGGTGGTGCAACACAAAACCCCGCCGACGCTGACGCTGAC 283881
Qy        458 CTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTC 517
Db     283880 CCTGCCGCGGCCCACCGTGGTGACCGGGCTGCGGCTGGCCGCCAGCCGGTCGATGTTGCC 283821
Qy        518 GGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGCAGGTCCACCTGCAGCA 577
Db     283820 GGCCCATCCGACGGTGGTGGCCATCAACCTGGGCGACGGCCCCCAGGTCCGACAACTGCA 283761
Qy        578 G 578
Db     283760 G 283760

RESULT 16
US-09-522-714-21
; Sequence 21, Application US/09522714
; Patent No. 6563020
; GENERAL INFORMATION:
;  APPLICANT: Simmons, Carl R.
;  APPLICANT: Yalpani, Nasser
;  TITLE OF INVENTION: Maize Chitinases and Their Use in
;  TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants
;  FILE REFERENCE: 1100
;  CURRENT APPLICATION NUMBER: US/09/522,714
;  CURRENT FILING DATE: 2000-03-10
;  EARLIER APPLICATION NUMBER: 60/125,915
;  EARLIER FILING DATE: 1999-03-24
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
;   LENGTH: 1048
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (42)...(854)
US-09-522-714-21

  Query Match             3.8%;  Score 37.6;  DB 4;  Length 1048;
  Best Local Similarity   42.6%;  Pred. No. 2.2;
  Matches  196;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy         83 CTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCC 142
Db        326 CCGCGACGCCTTCCTCGCCGCCGCCGGCTACTACCCGGCGTTCGGCCGCACCGGCACCGT 385
Qy        143 TGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCAT 202
Db        386 CGACGACTCCAAGCGCGAGATCGCCGCCTTCTTCGGCAACGCCAACCACGAGACCATAAA 445
Qy        203 AACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAA 262
Db        446 GTTCTGCTACATCAACGAGATCGACGGGCCGAGCAAGAACTACTGCGACCGGAACAACAC 505
Qy        263 CATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGT 322
Db        506 GCAGTGGCCGTGCCAGGCGGGGAAGGGGTACTACGGCCGCGGCCCGCTGCAGATCTCCTG 565
Qy        323 CACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGG 382
Db        566 GAACTTCAACTACGGGCCCGCGGGGCAGAGCATCGGCTTCGACGGGCTGGGCGACCCCGA 625
Qy        383 TGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTC 442
Db        626 CGCGGTGGCGCGCAGCGCCGTGCTCGCGTTCCGCTCCGCGCTCTGGTACTGGATGAACAA 685
Qy        443 TGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGAT 502
Db        686 CGTGCACGGGGCCATCGTCTCCGGCCAGGGCTTCGGCGCCACCATCCGGGCCATCAACGG 745
Qy        503 CTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGC 542
Db        746 CGCGCTCGAGTGCGACGGCAAGAACCCCAACTCCGTCAAC 785

RESULT 17
US-08-123-161A-11
; Sequence 11, Application US/08123161A
; Patent No. 5449616
;  GENERAL INFORMATION:
;    APPLICANT:  Campbell, Kevin P.
;    APPLICANT:  Roberds, Steven L.
;    APPLICANT:  Anderson, Richard D.
;    APPLICANT:  Ibraghimov, Oxana B.
;    APPLICANT:  Yang, Bin
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  USA
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/123,161A
;      FILING DATE:  16-SEP-93
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/946,234
;      FILING DATE:  14-SEP-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  UIRF89-11A4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1396 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..1164
US-08-123-161A-11

  Query Match             3.8%;  Score 37.6;  DB 1;  Length 1396;
  Best Local Similarity   65.5%;  Pred. No. 2.5;
  Matches   55;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        909 GCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTAGCTGCTCTTGCCA 968
```

```
Db       1309 GGGTGGGGTGAGAGTGTGTGGAGTAAGGACATTCAGAATAAATATCTGCTGCTCTGCTCA 1368

Qy        969 CAAAAAAAAAAAAAAAAAAAAAA 992
              |||||||||||||||||||||||
Db       1369 CCAAAAAAAAAAAAAAAAAAAAA 1392


RESULT 18
US-08-483-278-11
; Sequence 11, Application US/08483278
; Patent No. 5686073
;  GENERAL INFORMATION:
;     APPLICANT:  Campbell, Kevin P.
;     APPLICANT:  Ibraghimov, Oxana B.
;     APPLICANT:  Ervasti, James M.
;     APPLICANT:  Leveille, Cynthia J.
;     TITLE OF INVENTION:  NUCLEIC ACID ENCODING DYSTROPHIN-ASSOCIATED
;     TITLE OF INVENTION:  PROTEIN
;     NUMBER OF SEQUENCES:  15
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Kevin M. Farrell, P.C.
;       STREET:  P.O. Box 999
;       CITY:  York Harbor
;       STATE:  ME
;       COUNTRY:  USA
;       ZIP:  03911
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/483,278
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/123,161
;       FILING DATE:  16-SEP-93
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Farrell, Kevin M.
;       REGISTRATION NUMBER:  35,505
;       REFERENCE/DOCKET NUMBER:  UIRF89-11A5
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (207) 363-0558
;       TELEFAX:  (207) 363-0528
;   INFORMATION FOR SEQ ID NO:  11:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1396 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  4..1164
US-08-483-278-11

  Query Match             3.8%;  Score 37.6;  DB 1;  Length 1396;
  Best Local Similarity   65.5%;  Pred. No. 2.5;
  Matches   55;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        909 GCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCA 968
              |||  ||   || | ||||||     ||    || |   | |  |||||||  ||
Db       1309 GGGTGGGGTGAGAGTGTGTGGAGTAAGGACATTCAGAATAAATATCTGCTGCTCTGCTCA 1368

Qy        969 CAAAAAAAAAAAAAAAAAAAAAA 992
              |||||||||||||||||||||||
Db       1369 CCAAAAAAAAAAAAAAAAAAAAA 1392


RESULT 19
US-09-252-991A-5567/c
; Sequence 5567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 5567
;   LENGTH: 1992
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5567

  Query Match             3.8%;  Score 37.6;  DB 4;  Length 1992;
  Best Local Similarity   50.6%;  Pred. No. 2.8;
  Matches   91;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy        515 CTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCA 574
              ||||   |||| |     ||| | |  |||| ||   |  | | | ||  |    | ||
Db       1569 CTCGTCGCGCCCGCGCGCGTCCATGAACAGGATCAGGGCTGCGGTGTAGAGCGGGTAGAA 1510

Qy        575 GCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGA 634
              |   || ||      |||||  |  ||   | ||| ||    |||||       | |||
Db       1509 GAGCAGTCCGGCGACCAGGCTCCAGGCGCCCATCTGCTGGGGGCCGACGGCTTCGTCGAT 1450

Qy        635 CTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGT 694
              ||| | |||| ||  |  | | |  ||   | | | |  ||| |||| |  ||  |||||
Db       1449 CTGCTGCTGGAGCAGGCTTTCGAGGAATATCCAGGGCAGGCAGAGCGGCACCAGGGTGGT 1390


RESULT 20
US-08-997-080-179/c
; Sequence 179, Application US/08997080
; Patent No. 5968524
;  GENERAL INFORMATION:
;     APPLICANT:  WATSON, JAMES D.
;     APPLICANT:  TAN, PAUL L.J.
;     TITLE OF INVENTION:  METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
;     NUMBER OF SEQUENCES:  194
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Law Offices of Ann W. Speckman
;       STREET:  2601 Elliott Avenue, Suite 4185
;       CITY:  Seattle
;       STATE:  WA
;       COUNTRY:  USA
;       ZIP:  98121
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/997,080
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sleath, Janet
;       REGISTRATION NUMBER:  37,007
;       REFERENCE/DOCKET NUMBER:  11000.1007
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-269-0565
;       TELEFAX:  206-269-0563
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  179:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-997-080-179

  Query Match             3.8%;  Score 37.4;  DB 2;  Length 520;
  Best Local Similarity   48.0%;  Pred. No. 1.9;
  Matches  107;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
              | |||| |   |  | | |    || || |  | ||| | | | | | | |
Db        324 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 265

Qy        434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
              |||   ||| ||| ||||| |  | | |  ||  | ||| ||  | | |   ||
Db        264 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 205

Qy        494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
              |||  |  |  |    |||||||||          |  || | || ||||||||
Db        204 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 145

Qy        554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||| | ||  |    |  |  ||  |  |  ||| || |
Db        144 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 102


RESULT 21
US-08-997-362-179/c
; Sequence 179, Application US/08997362
; Patent No. 5985287
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Hiyama, Jun
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Scott, Linda
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
;    NUMBER OF SEQUENCES:  194
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,362
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. Patent Application No. 5985287 08/873,970
;      FILING DATE:  June 12, 1997
;      APPLICATION NUMBER:  U.S. Patent Application No. 5985287 08/705,347
;      FILING DATE:  August 29, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-997-362-179

  Query Match             3.8%;  Score 37.4;  DB 2;  Length 520;
  Best Local Similarity   48.0%;  Pred. No. 1.9;
  Matches  107;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
              | |||| |   |  | | |    || || |  | ||| | | | | | | |
Db        324 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 265

Qy        434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
              |||   ||| ||| ||||| |  | | |  ||  | ||| ||  | | |   ||
Db        264 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 205

Qy        494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
              |||  |  |  |    |||||||||          |  || | || ||||||||
Db        204 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 145

Qy        554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||| | ||  |    |  |  ||  |  |  ||| || |
Db        144 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 102


RESULT 22
US-09-095-855-179/c
; Sequence 179, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  Compounds and Methods for
;    TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/095,855
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/705,347
;      FILING DATE:  29-AUG-1996
;      APPLICATION NUMBER:  08/873,970
;      FILING DATE:  12-JUN-1997
;      APPLICATION NUMBER:  08/997,362
;      FILING DATE:  23-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sleath, Janet
;      REGISTRATION NUMBER:  37,007
;      REFERENCE/DOCKET NUMBER:  11000.1002c3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-269-0565
;      TELEFAX:  206-269-0563
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  179:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  520 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-09-095-855-179

  Query Match             3.8%;  Score 37.4;  DB 3;  Length 520;
  Best Local Similarity   48.0%;  Pred. No. 1.9;
  Matches  107;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
              | |||| |    |  | | |    || || |  | |||  | |  | |  | |  |
Db        324 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 265

Qy        434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
              |||   ||| ||| |||||  |   | | |  ||  |  |||  ||   | | |   ||
Db        264 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 205

Qy        494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
                ||| | |  |  |  |||||||||        |   || | || ||||||||
Db        204 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 145

Qy        554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||| | || |    |  |  | |  ||  | |  |||| | |
Db        144 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 102


RESULT 23
US-09-324-542-179/c
; Sequence 179, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Tan, Paul L.J.
;  APPLICANT: Prestidge, Ross
;  TITLE OF INVENTION: Methods and Compounds for the Treatment
;  TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
;  FILE REFERENCE: 11000.1007c1
;  CURRENT APPLICATION NUMBER: US/09/324,542
;  CURRENT FILING DATE: 1999-06-02
;  EARLIER APPLICATION NUMBER: US 08/997,080
;  EARLIER FILING DATE: 1997-12-23
;  NUMBER OF SEQ ID NOS: 194
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 179
;   LENGTH: 520
;   TYPE: DNA
;   ORGANISM: Mycobacterium vaccae
US-09-324-542-179

  Query Match             3.8%;  Score 37.4;  DB 4;  Length 520;
  Best Local Similarity   48.0%;  Pred. No. 1.9;
  Matches  107;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
              | |||| |    |  | | |    || || |  | |||  | |  | |  | |  |
Db        324 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 265

Qy        434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
              |||   ||| ||| |||||  |   | | |  ||  |  |||  ||   | | |   ||
Db        264 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 205

Qy        494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
                ||| | |  |  |  |||||||||        |   || | || ||||||||
Db        204 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 145

Qy        554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||| | || |    |  |  | |  ||  | |  |||| | |
Db        144 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 102


RESULT 24
US-09-205-426-179/c
; Sequence 179, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Tan, Paul L. J.
;  TITLE OF INVENTION: Compounds and Methods for Treatment and
;  TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
;  FILE REFERENCE: 11000.1002c4
;  CURRENT APPLICATION NUMBER: US/09/205,426
;  CURRENT FILING DATE: 1998-12-04
;  EARLIER APPLICATION NUMBER: 09/095,855
;  EARLIER FILING DATE: 1998-06-11
;  EARLIER APPLICATION NUMBER: 08/997,362
;  EARLIER FILING DATE: 1997-12-23
;  EARLIER APPLICATION NUMBER: 08/873,970
;  EARLIER FILING DATE: 1997-06-12
;  EARLIER APPLICATION NUMBER: 08/705,347
;  EARLIER FILING DATE: 1996-08-29
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 179
;   LENGTH: 520
;   TYPE: DNA
;   ORGANISM: Mycobacterium vaccae
US-09-205-426-179

  Query Match             3.8%;  Score 37.4;  DB 4;  Length 520;
  Best Local Similarity   48.0%;  Pred. No. 1.9;
  Matches  107;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
              | |||| |    |  | | |    || || |  | |||  | |  | |  | |  |
Db        324 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 265

Qy        434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
              |||   ||| ||| |||||  |   | | |  ||  |  |||  ||   | | |   ||
Db        264 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 205

Qy        494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
                ||| | |  |  |  |||||||||        |   || | || ||||||||
Db        204 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 145

Qy        554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
              ||| | || |    |  |  | |  ||  | |  |||| | |
Db        144 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 102


RESULT 25
US-09-095-855-202/c
; Sequence 202, Application US/09095855
; Patent No. 6160093
;  GENERAL INFORMATION:
;    APPLICANT:  Tan, Paul
;    APPLICANT:  Visser, Elizabeth
;    APPLICANT:  Skinner, Margot
;    APPLICANT:  Prestidge, Ross
;    TITLE OF INVENTION:  Compounds and Methods for
;    TITLE OF INVENTION:  Treatment and Diagnosis of Mycobacterial Infections
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Ann W. Speckman
;      STREET:  2601 Elliott Avenue, Suite 4185
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/095,855
;      FILING DATE:
```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 202:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 570 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-095-855-202

Query Match 3.8%; Score 37.4; DB 3; Length 570;
Best Local Similarity 48.0%; Pred. No. 2;
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
Db 333 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 274

Qy 434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
Db 273 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 214

Qy 494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
Db 213 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 154

Qy 554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
Db 153 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 111

RESULT 26
US-09-205-426-202/c
; Sequence 202, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 202
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Mycobacterium vaccae
US-09-205-426-202

Query Match 3.8%; Score 37.4; DB 4; Length 570;
Best Local Similarity 48.0%; Pred. No. 2;
Matches 107; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 374 CACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAA 433
Db 333 CTCCTCGGTGAGCTGCGGGATCGACACCCGGATGATGTTGCCGTCGTTGGTCGGATTGAC 274

Qy 434 GCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGT 493
Db 273 GCCGAGGTCGGAGTTGCGGATCGCATCCTCGATGAGGCGCAGCTGGCTCGCCTCGTAGGG 214

Qy 494 GGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAA 553
Db 213 CTTGATCACCACCATGCGCGCCTCGGGCACGTTGATGCTGGACAGCTGCGTGATCGGGGT 154

Qy 554 GGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAG 596
Db 153 GGAGGCGCCGTAGTAGTCGATGTTGATCCGGTTGAACATGCCG 111

RESULT 27
US-09-319-892-3
; Sequence 3, Application US/09319892
; Patent No. 6177616
; GENERAL INFORMATION:
; APPLICANT: BARTSCH, Klaus
; APPLICANT: KRIETE, Guido
; APPLICANT: BROER, Inge
; APPLICANT: PUHLER, Alfred
; TITLE OF INVENTION: NOVEL GENES CODING FOR AMINO ACID DEACETYLASES WITH
; TITLE OF INVENTION: SPECIFICITY FOR N-ACETYL-L-PHOSPHINOTHRICIN, THEIR
; TITLE OF INVENTION: ISOLATION AND THEIR USE
; FILE REFERENCE: 514412-2005
; CURRENT APPLICATION NUMBER: US/09/319,892
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/EP97/06755
; EARLIER FILING DATE: 1997-12-03
; EARLIER APPLICATION NUMBER: 19652284.6
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Stenotrophomonas maltophilia
US-09-319-892-3

Query Match 3.7%; Score 36.6; DB 3; Length 1273;
Best Local Similarity 45.0%; Pred. No. 4.3;
Matches 138; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

Qy 431 CAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAG 490
Db 735 CGAGGCCACGAACGTGGTGCCCGACAGCGTGGAGCTGCAGGGCACGGTGCGCACCTTCAC 794

Qy 491 GGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGG 550
Db 795 GCTGGAGGTGCTGGACCTGATCGAGCGGCGCATGAAGGCCCTGGCCGAGAGCATCTGCGC 854

Qy 551 GAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTC 610
Db 855 GGCGCATGACACGCGCTGCGAGTTCGAGTTCGTGCGCAACTACCCGCCCACCATCAACTC 914

Qy 611 CTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGT 670
Db 915 CGCCCCGGAGGCCGAGTTCGCACGCCGCGTCATGGCCGAGGTCGTGGGCGAGGCCAACGT 974

Qy 671 CCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCA 730
Db 975 GCTGCCCCAGGAGCCGTCCATGGGCGCCGAGGACTTCGCCTTCATGCTGCTGGAAAAGCC 1034

```
Qy        731 GGGTCCC 737
              ||  ||
Db       1035 CGGCGCC 1041
```

RESULT 28
US-09-727-238-3
; Sequence 3, Application US/09727238
; Patent No. 6555733
; GENERAL INFORMATION:
; APPLICANT: Bartsch, et al.
; TITLE OF INVENTION: No. 6555733el genes coding for amino acid deacetylases with speci
; TITLE OF INVENTION: acetyl-L-phosphinothricin, their isolation and their use.
; FILE REFERENCE: 514412A-2005.1
; CURRENT APPLICATION NUMBER: US/09/727,238
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/319,892
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: PCT/EP97/06755
; PRIOR FILING DATE: 1997-12-03
; PRIOR APPLICATION NUMBER: 196 52 284.6
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Comamonas acidovorans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1206)
; OTHER INFORMATION: coding sequence of the deacetylase gene from comamonas acidovoran
US-09-727-238-3

Query Match 3.7%; Score 36.6; DB 4; Length 1273;
Best Local Similarity 45.0%; Pred. No. 4.3;
Matches 138; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```
Qy        431 CAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAG 490
              | || |   |   | | ||| | || | |     ||||||| ||  | |     |  ||
Db        735 CGAGGCCACGAACGTGGTGCCCGACAGCGTGGAGCTGCAGGGCACGGTGCGCACCTTCAC 794

Qy        491 GGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGG 550
              | ||||| || |   || |       | |||     |  |  || | ||| |   ||
Db        795 GCTGGAGGTGCTGGACCTGATCGAGCGGCGCATGAAGGCCCTGGCCGAGAGCATCTGCGC 854

Qy        551 GAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTC 610
              |  | |||     |  |  |  |    ||       || ||   | |||  ||| |  |||
Db        855 GGCGCATGACACGCGCTGCGAGTTCGAGTTCGTGCGCAACTACCCGCCCACCATCAACTC 914

Qy        611 CTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAATGT 670
              |  || |  |  | |||    |   |  ||  |  | |||  |     | | |||||  ||
Db        915 CGCCCCGGAGGCCGAGTTCGCACGCCGCGTCATGGCCGAGGTCGTGGGCGAGGCCAACGT 974

Qy        671 CCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCA 730
              | || |  |  |  | |||   ||  |   | |  | | ||  |||   ||
Db        975 GCTGCCCCAGGAGCCGTCCATGGGCGCCGAGGACTTCGCCTTCATGCTGCTGGAAAAGCC 1034

Qy        731 GGGTCCC 737
              ||  ||
Db       1035 CGGCGCC 1041
```

RESULT 29
US-09-252-991A-11102
; Sequence 11102, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11102
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11102

Query Match 3.7%; Score 36.4; DB 4; Length 801;
Best Local Similarity 50.0%; Pred. No. 4.1;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||  | |  |  ||    | | |    |  | || || |   | |||  ||| |
Db        279 CCAAGGCTTTCGCCGAGCGCCTCGACAAGCTGTGCCGGATCAGCGTCAAGGAAGCAGAGG 338

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||| | |  ||| ||     || |||  |  || || | || || ||  | |||| |||||
Db        339 ACGGCGACATCCTGCGTCCCGGCCTGGCGCTGCTGGCGCCGGGCGGCAAGCAGATGATGG 398

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              |||  |  ||  |  ||   | ||  |  |  |  |  ||  ||   ||
Db        399 TCGACGGCCGCGGCACGGTGCGCATCCTCCCCGGCGACGAACGCCTGAACTACAAGCCCT 458

Qy        781 GC 782
              ||
Db        459 GC 460
```

RESULT 30
US-09-252-991A-11158
; Sequence 11158, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11158
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11158

Query Match 3.7%; Score 36.4; DB 4; Length 1116;
Best Local Similarity 50.0%; Pred. No. 4.6;
Matches 91; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||  | |  |  ||    | | |    |  | || || |   | |||  ||| |
Db        683 CCAAGGCTTTCGCCGAGCGCCTCGACAAGCTGTGCCGGATCAGCGTCAAGGAAGCAGAGG 742

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||| | |  ||| ||     || |||  |  || || | || || ||  | |||| |||||
Db        743 ACGGCGACATCCTGCGTCCCGGCCTGGCGCTGCTGGCGCCGGGCGGCAAGCAGATGATGG 802

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              |||  |  ||  |  ||   | ||  |  |  |  |  ||  ||   ||
Db        803 TCGACGGCCGCGGCACGGTGCGCATCCTCCCCGGCGACGAACGCCTGAACTACAAGCCCT 862

Qy        781 GC 782
              ||
```

```
Db        863 GC 864


RESULT 31
US-09-252-991A-11094/c
; Sequence 11094, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 11094
;   LENGTH: 1584
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11094

  Query Match             3.7%;  Score 36.4;  DB 4;  Length 1584;
  Best Local Similarity   50.0%;  Pred. No. 5.2;
  Matches   91;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||    | ||  |  ||    | | |    |  | || || |   | |||  ||| |
Db        839 CCAAGGCTTTCGCCGAGCGCCTCGACAAGCTGTGCCGGATCAGCGTCAAGGAAGCAGAGG 780

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||| | |  ||| ||    || |||  |  || || | || || ||  | |||| |||||
Db        779 ACGGCGACATCCTGCGTCCCGGCCTGGCGCTGCTGGCGCCGGGCGGCAAGCAGATGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||| | |  ||  |  ||  | || |  |  | |  |  || ||  |  ||
Db        719 TCGACGGCCGCGGCACGGTGCGCATCCTCCCCGGCGACGAACGCCTGAACTACAAGCCCT 660

Qy        781 GC 782
              ||
Db        659 GC 658


RESULT 32
US-09-211-930-12
; Sequence 12, Application US/09211930
; Patent No. 5962265
; GENERAL INFORMATION:
;  APPLICANT: Tyrell E. No. 5962265ris
;  APPLICANT: William Craig Moore
;  APPLICANT: David Shay Silberstein
;  TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION
;  TITLE OF INVENTION: SERINE/THREONINE KINASE
;  FILE REFERENCE: PHM.70296
;  CURRENT APPLICATION NUMBER: US/09/211,930
;  CURRENT FILING DATE: 1998-12-15
;  EARLIER APPLICATION NUMBER: GB 9726851.0
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 2028
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-211-930-12

  Query Match             3.7%;  Score 36.4;  DB 2;  Length 2028;
  Best Local Similarity   79.6%;  Pred. No. 5.7;
  Matches   43;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy        939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||  |||||  |||  ||  ||||  |  | ||||||||||||||||||||||
Db       1950 AAAACAAGTGTTTTTAAGAAGCTCCCGGAAAAAAAAAAAAAAAAAAAAAAAAAA 2003


RESULT 33
US-09-340-993-12
; Sequence 12, Application US/09340993
; Patent No. 6034228
; GENERAL INFORMATION:
;  APPLICANT: Tyrell E. No. 6034228ris
;  APPLICANT: William Craig Moore
;  APPLICANT: David Shay Silberstein
;  TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
;  FILE REFERENCE: PHM.70296.N1
;  CURRENT APPLICATION NUMBER: US/09/340,993
;  CURRENT FILING DATE: 1999-06-25
;  EARLIER APPLICATION NUMBER: GB 9726851.0 & US 09/211,930
;  EARLIER FILING DATE: 1997-12-19 & 1998-12-15
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 2028
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-340-993-12

  Query Match             3.7%;  Score 36.4;  DB 3;  Length 2028;
  Best Local Similarity   79.6%;  Pred. No. 5.7;
  Matches   43;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy        939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||  |||||  |||  ||  ||||  |  | ||||||||||||||||||||||
Db       1950 AAAACAAGTGTTTTTAAGAAGCTCCCGGAAAAAAAAAAAAAAAAAAAAAAAAAA 2003


RESULT 34
US-09-468-442-12
; Sequence 12, Application US/09468442
; Patent No. 6300098
; GENERAL INFORMATION:
;  APPLICANT: Tyrell E. No. 6300098ris
;  APPLICANT: William Craig Moore
;  APPLICANT: David Shay Silberstein
;  TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
;  FILE REFERENCE: PHM.70296.N1
;  CURRENT APPLICATION NUMBER: US/09/468,442
;  CURRENT FILING DATE: 1999-12-21
;  EARLIER APPLICATION NUMBER: US 09/340,993
;  EARLIER FILING DATE: 1999-06-25
;  EARLIER APPLICATION NUMBER: GB 9726851.0
;  EARLIER FILING DATE: 1997-12-19
;  EARLIER APPLICATION NUMBER: US 09/211,930
;  EARLIER FILING DATE: 1998-12-15
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
;   LENGTH: 2028
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-468-442-12

  Query Match             3.7%;  Score 36.4;  DB 4;  Length 2028;
  Best Local Similarity   79.6%;  Pred. No. 5.7;
  Matches   43;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy        939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||  |||||  |||  ||  ||||  |  | ||||||||||||||||||||||
Db       1950 AAAACAAGTGTTTTTAAGAAGCTCCCGGAAAAAAAAAAAAAAAAAAAAAAAAAA 2003


RESULT 35
US-09-252-991A-5589/c
```

; Sequence 5589, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5589
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5589

Query Match 3.6%; Score 36.2; DB 4; Length 570;
Best Local Similarity 45.6%; Pred. No. 4;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy        370 CCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGT 429
Db        513 CCGGCACCAGCCCGGCCAGCGGTCCCTGCAACTGGCTATAGCCGGCGTAGAGCCAGGGGA 454

Qy        430 CCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCA 489
Db        453 AACCGGTAAGGAACCAACTGCGGAACAGCTCCAGTGCCAGCCACAGGGCGGCGAACGCCA 394

Qy        490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
Db        393 GGGCGTCGCCCAGCGGCGCGTTGTCGCGGCGCAGGCAGCGCGCCCACAGCCAGGCCGGCA 334

Qy        550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
Db        333 GGGCGAAGAAGAAGGCGACGCCGGCGGTGAAGCCGAGCATGAGGAGGCTTGCCAGCGGCA 274

Qy        610 CCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAG 650
Db        273 CCGAGGCGGCGCCGTAGTCATGGATACTGACGTAGATCCAG 233
```

RESULT 36
US-09-252-991A-5553/c
; Sequence 5553, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5553
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5553

Query Match 3.6%; Score 36.2; DB 4; Length 1545;
Best Local Similarity 45.6%; Pred. No. 5.8;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy        370 CCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGT 429
Db        502 CCGGCACCAGCCCGGCCAGCGGTCCCTGCAACTGGCTATAGCCGGCGTAGAGCCAGGGGA 443

Qy        430 CCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCA 489
Db        442 AACCGGTAAGGAACCAACTGCGGAACAGCTCCAGTGCCAGCCACAGGGCGGCGAACGCCA 383

Qy        490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
Db        382 GGGCGTCGCCCAGCGGCGCGTTGTCGCGGCGCAGGCAGCGCGCCCACAGCCAGGCCGGCA 323

Qy        550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
Db        322 GGGCGAAGAAGAAGGCGACGCCGGCGGTGAAGCCGAGCATGAGGAGGCTTGCCAGCGGCA 263

Qy        610 CCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAG 650
Db        262 CCGAGGCGGCGCCGTAGTCATGGATACTGACGTAGATCCAG 222
```

RESULT 37
US-09-252-991A-5670
; Sequence 5670, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5670
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5670

Query Match 3.6%; Score 36.2; DB 4; Length 1914;
Best Local Similarity 45.6%; Pred. No. 6.3;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
Qy        370 CCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGT 429
Db        987 CCGGCACCAGCCCGGCCAGCGGTCCCTGCAACTGGCTATAGCCGGCGTAGAGCCAGGGGA 1046

Qy        430 CCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCA 489
Db       1047 AACCGGTAAGGAACCAACTGCGGAACAGCTCCAGTGCCAGCCACAGGGCGGCGAACGCCA 1106

Qy        490 GGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCG 549
Db       1107 GGGCGTCGCCCAGCGGCGCGTTGTCGCGGCGCAGGCAGCGCGCCCACAGCCAGGCCGGCA 1166

Qy        550 GGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCT 609
Db       1167 GGGCGAAGAAGAAGGCGACGCCGGCGGTGAAGCCGAGCATGAGGAGGCTTGCCAGCGGCA 1226

Qy        610 CCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAG 650
Db       1227 CCGAGGCGGCGCCGTAGTCATGGATACTGACGTAGATCCAG 1267
```

RESULT 38
US-09-245-041-18
; Sequence 18, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.

```
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
;  TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
;  FILE REFERENCE: 7853-136
;  CURRENT APPLICATION NUMBER: US/09/245,041
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/093,630
;  EARLIER FILING DATE: 1998-07-21
;  EARLIER APPLICATION NUMBER: 60/104,978
;  EARLIER FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
;   LENGTH: 2625
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-245-041-18

  Query Match             3.6%;  Score 36;  DB 3;  Length 2625;
  Best Local Similarity   67.1%;  Pred. No. 7.9;
  Matches   51;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy        917 TGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAA 976
Db       2528 TGTTCCTACTGTCCTTTGAAATAAAGTGAAACATCCTTTGCTGCTCTGTAAAAAAAAAAA 2587

Qy        977 AAAAAAAAAAAAAAAA 992
Db       2588 AAAAAAAAAAAAAAAA 2603

RESULT 39
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             3.6%;  Score 36;  DB 3;  Length 4403765;
  Best Local Similarity   47.9%;  Pred. No. 69;
  Matches  134;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    1;

Qy        239 CTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGA 298
Db    3773464 CGACGCGCCGTATTGGGCCAAGTACCGCGAGGATGCCGCGAAGTTCGGGCTGACCGAGGC 3773523

Qy        299 GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTT 358
Db    3773524 GATCGCCGCCTACAGCACCCGGCTGAAGATCACCCCGACCCGGGTGTGGACGACGCCCAC 3773583

Qy        359 CTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTG 418
Db    3773584 GGGCTGAGCGGGCTGGCCCCCGCTCGCCGCCAGAGTGAAATCCACGACGCGTTTGCGGCG 3773643

Qy        419 GGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG 478
Db    3773644 TGTCGCGTCGCCCGTTTCACTGTCGGCGCAGAGGTTCACCGGAAGTC-GCGCGAGCGCGC 3773702

Qy        479 GGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCG 518
Db    3773703 GCCGACCGCCAGGGTGAGGCGGCCCATCCGTTCGGCGACG 3773742

RESULT 40
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             3.6%;  Score 36;  DB 3;  Length 4411529;
  Best Local Similarity   47.9%;  Pred. No. 69;
  Matches  134;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    1;

Qy        239 CTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGA 298
Db    3781282 CGACGCGCCGTATTGGGCCAAGTACCGCGAGGATGCCGCGAAGTTCGGGCTGACCGAGGC 3781341

Qy        299 GCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTT 358
Db    3781342 GATCGCCGCCTACAGCACCCGGCTGAAGATCACCCCGACCCGGGTGTGGACGACGCCCAC 3781401

Qy        359 CTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTG 418
Db    3781402 GGGCTGAGCGGGCTGGCCCCCGCTCGCCGCCAGAGTGAAATCCACGACGCGTTTGCGGCG 3781461

Qy        419 GGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGG 478
Db    3781462 TGTCGCGTCGCCCGTTTCACTGTCGGCGCAGAGGTTCACCGGAAGTC-GCGCGAGCGCGC 3781520

Qy        479 GGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCG 518
Db    3781521 GCCGACCGCCAGGGTGAGGCGGCCCATCCGTTCGGCGACG 3781560

RESULT 41
US-08-258-261B-6/c
; Sequence 6, Application US/08258261B
; Patent No. 5639949
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,261B
;      FILING DATE:  08-JUN-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 42
US-08-456-837-6/c
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/456,837
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-456-837-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 43
US-08-457-342-6/c
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
```

```
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,342
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-457-342-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 44
US-08-457-646A-6/c
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,646A
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-457-646A-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 45
US-08-458-076A-6/c
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
```

```
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/458,076A
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-458-076A-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |  ||       |||||| |||| || | || |   |||
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | |||||  |  || |||  | ||| |  ||| |   |      ||  | | |
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                 |||   | || | |       |  ||||
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 46
US-08-764-233A-4/c
; Sequence 4, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1
US-08-764-233A-4

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |  ||       |||||| |||| || | || |   |||
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | |||||  |  || |||  | ||| |  ||| |   |      ||  | | |
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                 |||   | || | |       |  ||||
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 47
US-08-457-335A-6/c
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, PHillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
```

```
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,335A
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-457-335A-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |   ||      ||||||| |||| || | || |   |||
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | ||||| |  || |||  | ||| |  ||| |   |     ||  | | |
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                 |||   | || | |      |  ||||
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 48
US-08-729-214-6/c
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |   ||      ||||||| |||| || | || |   |||
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | ||||| |  || |||  | ||| |  ||| |   |     ||  | | |
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                 |||   | || | |      |  ||||
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 49
US-09-028-934-6/c
; Sequence 6, Application US/09028934
; Patent No. 6117670
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Lam, Steven T.
;    APPLICANT:  Hammer, Philip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    APPLICANT:  Young, Thomas R.
;    TITLE OF INVENTION:  Pyrrolnitrin Biosynthesis Genes and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6117670artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  NC
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/09/028,934
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/729,214
;       FILING DATE:  09-OCT-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/258,261
;       FILING DATE:  08-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meigs, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  CGC1506/CIP7
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8587
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-09-028-934-6

  Query Match             3.6%;  Score 35.8;  DB 3;  Length 28958;
  Best Local Similarity   52.3%;  Pred. No. 21;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |   ||       ||||| |||| || | || |   |||
Db       4498 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 4439

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | ||||| |  || |||  | ||| |  ||| |    |       ||  | | |
Db       4438 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 4379

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                 |||   | || | |       |  ||||
Db       4378 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 4348


RESULT 50
US-08-764-233A-1/c
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homologous
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
;      OTHER INFORMATION:  /product= "SorA"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs that
;      OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;      OTHER INFORMATION:  compounds."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  942..7115
;      OTHER INFORMATION:  /product= "Module 1 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7203..12884
;      OTHER INFORMATION:  /product= "Module 2 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  13455..19616
;      OTHER INFORMATION:  /product= "Module 3 of SorA"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19871..46318
;      OTHER INFORMATION:  /product= "SorB"
;      OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS genes
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19870..24556
;      OTHER INFORMATION:  /product= "Module 1 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  24638..30820
;      OTHER INFORMATION:  /product= "Module 2 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  30881..35446
;      OTHER INFORMATION:  /product= "Module 3 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  35528..40114
;      OTHER INFORMATION:  /product= "Module 4 of SorB"
;    FEATURE:
;      NAME/KEY:  misc_feature
```

```
;       LOCATION:  40190..46318
;       OTHER INFORMATION:  /product= "Module 5 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  46851..47891
;       OTHER INFORMATION:  /product= "SorM"
;       OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match             3.6%;  Score 35.8;  DB 1;  Length 49377;
  Best Local Similarity   52.3%;  Pred. No. 26;
  Matches   79;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         56 CCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTC 115
              | | | ||| |||| |  || |   ||      ||||||| |||| || | || |   |||
Db      20801 CAGGGCCATTGGGCGCGGTGAGGCCTTGGCTCCGGCCGTCCTGGTTGACGGCGGAGCCTC 20742

Qy        116 CAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGT 175
              ||||   | ||||| |  || |||  | ||| |  ||| |    |      ||  | | |
Db      20741 GAAGGACGGCGAGGACGGGATGACCGTTTTGGACGGCATCGGAGAGGCGCTCGAGCAGGA 20682

Qy        176 TTTCCCCAAAGGCCGCTGGGTGCTCATAACC 206
                |||   | || | |      |  ||||
Db      20681 GCATCCCGGCGCCCTCGGCCCAGCCCGAACC 20651


Search completed: February  9, 2004, 15:01:36
Job time : 177 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: February 9, 2004, 12:59:37 ; Search time 328 Seconds
(without alignments)
8164.154 Million cell updates/sec

Title: US-09-990-726-222
Perfect score: 992
Sequence: 1 ggcacgagccaggaactagg.....aaaaaaaaaaaaaaaaaaa 992

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	992	100.0	992	21	AAC58373	Human PRO809 nucle
2	992	100.0	992	22	AAF44176	Human PRO809 (UNQ4
3	992	100.0	992	24	ABK33610	cDNA encoding huma
4	992	100.0	992	25	ABX80267	Novel human secret
5	992	100.0	992	25	ABX80771	Human secreted/tra
6	992	100.0	992	25	ABX81154	Novel human secret
7	992	100.0	992	25	ABX90244	Human secreted/tra
8	992	100.0	992	25	ABX77855	Human PRO polynucl
9	992	100.0	992	25	ABX79451	Human sectreted/tr
10	992	100.0	992	25	ABX64090	cDNA encoding huma
11	992	100.0	992	25	ABX17054	Human PRO polynucl
12	991	99.9	991	21	AAZ65030	Membrane-bound pro
13	843	85.0	1047	22	AAH64779	Human secreted pro
14	790.2	79.7	935	22	AAH99131	Human EST-derived
15	773.6	78.0	1047	21	AAC74260	Human secreted pro
16	773.6	78.0	1047	25	ABT16794	Human secreted pro
17	773.6	78.0	1047	25	ABZ66993	Human secreted pro
18	501.6	50.6	502	25	ABZ36822	Human GENSET codin
19	501.6	50.6	724	22	AAS34076	Human cDNA encodin
20	292.8	29.5	32220	22	ABA20618	Human nervous syst
c 21	149.6	15.1	150	19	AAX12008	Human biallelic po
c 22	149.6	15.1	150	19	AAX12009	Human biallelic po
23	47.2	4.8	1272	24	ABX97141	Human NOV93a cDNA.
24	47	4.7	53226	25	ABQ76896	Human G-protein co
c 25	47	4.7	87350	18	AAX83003	Human WRN genomic
c 26	44.4	4.5	636	23	AAS68336	DNA encoding novel
27	44	4.4	249999	25	ABZ80229	Human tramdorin ge
28	43.4	4.4	560	24	ABT07673	Human breast cance
29	43	4.3	509	24	ABT07672	Human breast cance
c 30	42.8	4.3	143899	24	AAL38336	Genomic sequence e
c 31	41.8	4.2	752	21	AAF21712	Human breast and o
c 32	41.8	4.2	1417	21	AAC77293	Human ORFX ORF2848
33	41.8	4.2	2059	21	AAC77637	Human cancer assoc
c 34	41.8	4.2	2672	25	ABX63699	Human cDNA #699 di
c 35	41.8	4.2	3471	22	AAH15559	Human cDNA sequenc
c 36	41.8	4.2	3541	24	ABL53715	Human poly(ADP-rib
37	41	4.1	38736	24	ABQ99652	Human membrane spa
38	39.6	4.0	44100	21	ABN97975	Human retroviral s
39	39.4	4.0	276	24	ABN22420	Human ORFX polynuc
40	39.4	4.0	3285	20	AAX80672	Clone nn103_4 enco
41	39.4	4.0	3285	22	AAS59222	Human cDNA encodin
42	39.4	4.0	3285	24	ABA90891	Human polynucleoti
c 43	39.2	4.0	394	22	AAF65032	Novel human polynu
44	38.8	3.9	1639	21	AAC98202	Human colon cancer
45	38.8	3.9	2025	24	ABK33656	cDNA encoding huma
c 46	38.6	3.9	56737	24	ABS69895	Human hypoxanthine
47	38.4	3.9	185695	25	ABQ77405	Human THBS1 DNA.
c 48	38	3.8	294	23	ABV61090	Human prostate exp
49	38	3.8	349980	24	ABQ81848	Bifidobacterium lo
50	37.8	3.8	353	23	ABV18588	Human prostate exp
c 51	37.8	3.8	4403765	22	AAI99683	Mycobacterium tube
c 52	37.8	3.8	4411529	22	AAI99682	Mycobacterium tube
53	37.6	3.8	375	22	AAI88830	Human polynucleoti
54	37.6	3.8	861	24	ABA90348	Human polynucleoti
55	37.6	3.8	1048	21	AAA96232	cDNA encoding a ma
56	37.6	3.8	1396	16	AAQ98498	Human 50 kDa dystr
57	37.6	3.8	1396	18	AAT97324	Human (50 kDa) dys
58	37.6	3.8	1602	24	ABA90349	Human polynucleoti
59	37.6	3.8	1626	25	ABT31957	High-affinity phos
60	37.6	3.8	1626	25	ABT32039	High-affinity OsPH
61	37.6	3.8	3083	25	AAL51205	Human gene sequenc
62	37.6	3.8	3537	25	AAL51404	Human secreted pro
63	37.6	3.8	80959	25	AAL51405	Human secreted pro
c 64	37.4	3.8	520	20	AAZ11382	Nucleotide sequenc
c 65	37.4	3.8	520	24	ABL36288	M vaccae GV-41 cod
c 66	37.4	3.8	570	20	AAZ11394	Nucleotide sequenc
67	37.4	3.8	3675	21	AAF21797	Human breast and o
c 68	37.2	3.8	296	22	AAL21959	Human breast cance
69	37.2	3.8	450	19	AAV25967	Cysteine protease
70	37	3.7	697	23	ABL15733	Drosophila melanog
c 71	36.8	3.7	512	24	ABL87279	Human ovarian canc
c 72	36.8	3.7	545	24	ABQ58601	Human colon cancer
73	36.8	3.7	784	23	ABL08343	Drosophila melanog
74	36.8	3.7	2693	23	ABL11520	Drosophila melanog
75	36.8	3.7	2784	23	ABL08342	Drosophila melanog
76	36.6	3.7	1273	19	AAV45110	Comamonas sp. N-Ac
c 77	36.6	3.7	1571	24	ABL90445	Human polynucleoti
c 78	36.4	3.7	381	20	AAV88083	EST clone DN809.
c 79	36.4	3.7	381	20	AAV87989	EST clone FN251.
c 80	36.4	3.7	745	19	AAV82783	Clone dn809_5 isol
c 81	36.4	3.7	745	24	ABQ92020	Human polynucleoti

c 82	36.4	3.7	760	20	AAX61427	DNA encoding a hum
c 83	36.4	3.7	760	25	ACC50346	Human secreted pro
c 84	36.4	3.7	760	25	ABZ71191	Human secreted pro
85	36.4	3.7	809	21	AAC79824	Human secreted pro
86	36.4	3.7	2028	20	AAX81861	cDNA encoding a mu
87	36.4	3.7	2028	21	AAA08010	Mouse polynucleoti
88	36.4	3.7	2044	20	AAV84372	Human epidermoid c
89	36.2	3.6	295	22	AAS29052	cDNA encoding for
90	36.2	3.6	295	24	ABS68192	cDNA encoding huma
91	36.2	3.6	304	22	AAS29128	cDNA encoding for
92	36.2	3.6	304	24	ABS68268	cDNA encoding huma
c 93	36.2	3.6	359	22	AAL23694	Human breast cance
c 94	36.2	3.6	590	24	ABL59497	EST related to hum
95	36.2	3.6	1641	20	AAZ10823	Choline oxidase (C
96	36.2	3.6	2821	22	AAI66075	Triticum aestivum
97	36.2	3.6	2920	24	ABL35012	Rat cDNA isolated
98	36	3.6	541	20	AAX87415	Hepatocellular car
99	36	3.6	839	22	AAC86484	Maize Mac20 coding
100	36	3.6	1127	22	AAD07899	Human secreted pro
101	36	3.6	1128	22	AAD07835	Human secreted pro
102	36	3.6	1805	21	AAZ58260	Soybean histone de
c 103	36	3.6	1975	24	ABL90817	Human polynucleoti
104	36	3.6	2101	21	AAC93390	Human secreted pro
105	36	3.6	2101	22	AAF32713	Human secreted pro
106	36	3.6	2101	25	ABZ73609	Secreted protein-e
107	36	3.6	2101	25	ABZ67203	Human secreted pro
108	36	3.6	2625	21	AAZ91922	Human mahogany pro
109	36	3.6	2626	16	AAQ94582	Modified Taq DNA-p
110	36	3.6	10579	22	ABA20014	Human nervous syst
111	36	3.6	10579	22	AAK70045	Human immune/haema
112	36	3.6	21185	21	AAA63350	Streptomyces globi
113	36	3.6	28995	22	ABA20015	Human nervous syst
114	36	3.6	28995	22	AAK70046	Human immune/haema
c 115	36	3.6	28995	22	AAK79967	Human immune/haema
c 116	36	3.6	28995	22	AAK85213	Human immune/haema
117	36	3.6	63164	21	AAA63348	Streptomyces globi
118	36	3.6	4403765	22	AAI99683	Mycobacterium tube
119	36	3.6	4411529	22	AAI99682	Mycobacterium tube
120	35.8	3.6	879	24	ABN98766	Arabidopsis thalia
121	35.8	3.6	1132	21	AAC76120	Human ORFX ORF1675
122	35.8	3.6	1725	24	ABA95144	Human lysyl-oxidas
123	35.8	3.6	2328	24	AAI67786	Human lysyl oxidas
c 124	35.8	3.6	28598	17	AAT06769	Sorangium cellulos
c 125	35.8	3.6	28958	18	AAT89956	Sorangium cellulos
c 126	35.8	3.6	28958	21	AAA75299	DNA sequence of So
c 127	35.8	3.6	49377	19	AAV05287	The soraphen biosy
128	35.8	3.6	144460	21	AAZ93815	Olfactory receptor
c 129	35.8	3.6	1503841	24	ABT00010	Human neuregulin 1
c 130	35.8	3.6	1503841	24	ABT01503	Human neuregulin 1
c 131	35.8	3.6	1503900	22	AAK95240	Human neuregulin-1
c 132	35.8	3.6	1503900	22	AAK96733	Human neuregulin-1
133	35.6	3.6	366	21	AAZ92717	Human wild-type cy
134	35.6	3.6	387	21	AAZ92716	Human wild-type cy
c 135	35.6	3.6	500	21	AAC54277	Arabidopsis thalia
136	35.6	3.6	577	24	ABL66282	Lung cancer relate
137	35.6	3.6	577	24	ABL67527	Thyroid cancer rel
138	35.6	3.6	577	24	ABL67787	Oesophagus cancer
139	35.6	3.6	581	21	AAZ35658	Human cystatin E e
140	35.6	3.6	588	21	AAZ60085	Human cystatin E (
141	35.6	3.6	588	22	AAD20097	Human cystatin E (
142	35.6	3.6	588	24	AAL43452	Human cystatin E (
143	35.6	3.6	591	24	ABZ35213	Human gene express
144	35.6	3.6	598	18	AAT60608	Human cysteine pro
c 145	35.6	3.6	1536	23	AAS51510	Pseudomonas aerugi
146	35.6	3.6	2046	24	ABQ54507	Human ovarian anti
c 147	35.6	3.6	16759	24	ABL36297	Human lysosomal ac
c 148	35.4	3.6	294	23	ABV61758	Human prostate exp
c 149	35.4	3.6	356	23	ABV10788	Human prostate exp
c 150	35.4	3.6	361	23	ABV01619	Human prostate exp

ALIGNMENTS

RESULT 1
AAC58373
ID AAC58373 standard; cDNA; 992 BP.
XX
AC AAC58373;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO809 nucleotide sequence SEQ ID NO:22.
XX
KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumourigenesis; identification; cancer; cytostatic;
KW nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophagal disorder;
KW epithelial disorder; stromal disorder; blastocoelic disorder;
KW inflammatory disorder; immunologic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200053755-A2.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000WO-US00376.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99WO-US28313.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
PI Watanabe CK, Wood WI;
XX
DR WPI; 2000-572270/53.
DR P-PSDB; AAB24063.
XX
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer -
XX
PS Claim 50; Fig 13; 286pp; English.
XX
CC The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds maybe used to treat various conditions, including
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoelic disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.

```
XX
SQ   Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

  Query Match             100.0%;  Score 992;  DB 21;  Length 992;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-236;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 2
AAF44176
ID   AAF44176 standard; cDNA; 992 BP.
XX
AC   AAF44176;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO809 (UNQ464) nucleotide sequence SEQ ID NO:222.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   P-PSDB; AAB65214.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 2; Fig 150; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
```

```
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

  Query Match           100.0%;  Score 992;  DB 22;  Length 992;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-236;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 3
ABK33610
ID   ABK33610 standard; cDNA; 992 BP.
XX
AC   ABK33610;
XX
DT   08-MAY-2002  (first entry)
XX
DE   cDNA encoding human PRO protein, Seq ID No 149.
XX
KW   Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW   breast cancer; prostate tumour; rectal tumour; liver tumour;
KW   pericyte cell proliferation; chondrocyte cell proliferation;
KW   tumour necrosis factor-alpha; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200208288-A2.
XX
PD   31-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-US21066.
XX
PR   20-JUL-2000; 2000US-219556P.
PR   25-JUL-2000; 2000US-220585P.
PR   25-JUL-2000; 2000US-220605P.
PR   25-JUL-2000; 2000US-220607P.
PR   25-JUL-2000; 2000US-220624P.
PR   25-JUL-2000; 2000US-220638P.
PR   25-JUL-2000; 2000US-220664P.
PR   25-JUL-2000; 2000US-220666P.
PR   26-JUL-2000; 2000US-220893P.
PR   28-JUL-2000; 2000WO-US20710.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   15-SEP-2000; 2000US-000000P.
PR   10-NOV-2000; 2000WO-US30873.
PR   28-NOV-2000; 2000US-253646P.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000US-0747259.
PR   20-DEC-2000; 2000WO-US34956.
PR   28-FEB-2001; 2001WO-US06520.
PR   10-MAY-2001; 2001US-0854280.
PR   25-MAY-2001; 2001WO-US17092.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desnoyers L,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Smith V,  Stephan JF,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2002-172001/22.
DR   P-PSDB; AAU83666.
XX
PT   One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT   useful for treating a PRO related disorder and for diagnosing tumours
PT   such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT   tumour or liver tumour -
XX
PS   Claim 2; Figure 149; 359pp; English.
XX
CC   The invention relates to one hundred and twenty two nucleic acids
CC   encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
```

```
CC   encode human secreted proteins. The PRO nucleic acids, polypeptides,
CC   agonists and antagonists are useful for treating a PRO related disorder.
CC   The PRO polypeptides are useful for diagnosing tumours, especially lung
CC   cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
CC   liver tumour. The PRO polypeptides are useful for stimulating the
CC   proliferation of, or gene expression, in pericyte cells, for stimulating
CC   the proliferation or differentiation of chondrocyte cells, for
CC   stimulating the release of tumour necrosis factor-alpha from human blood,
CC   for stimulating or inhibiting the proliferation of normal human dermal
CC   fibroblast cells. The PRO polypeptide may also be used as molecular
CC   weight markers and for tissue typing. The PRO nucleic acids have
CC   applications in molecular biology, including use as hybridisation probes,
CC   and in chromosome and gene mapping. ABK33536-ABK33657 represent human
CC   PRO protein coding sequences of the invention.
XX
SQ   Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

  Query Match           100.0%;  Score 992;  DB 24;  Length 992;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-236;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 4
ABX80267
ID   ABX80267 standard; DNA; 992 BP.
XX
AC   ABX80267;
XX
DT   28-APR-2003  (first entry)
XX
DE   Novel human secreted or transmembrane protein PRO791 DNA.
XX
KW   Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
KW   cardiac insufficiency disorder; cancer; tumour; immune response;
KW   adrenal cortical capillary endothelial growth; c-fos induction;
KW   vascular endothelial growth factor inhibition; VEGF inhibition;
KW   endothelial cell growth inhibitor; T-lymphocytes stimulation;
KW   retinal neurons cell survival; rod photoreceptor cell survival;
KW   retinal disorder; retinitis pigmentosum; kidney disorder;
KW   mammalian kidney mesangial cell proliferation; Berger disease;
KW   dermatitis; herpetiformis; Crohn's disease; chondrocyte proliferation;
KW   chondrocyte redifferentiation; sports injury; arthritis; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2002132252-A1.
XX
PD   19-SEP-2002.
XX
PF   14-NOV-2001; 2001US-0990442.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   06-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   30-MAR-2000; 2000WO-US08439.
```

PR 15-MAY-2000; 2000WO-US13358.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.
PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 28-AUG-2001; 2001US-0941992.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI; 2003-247083/24.
DR P-PSDB; ABU59108.
XX
PT Novel isolated PRO polypeptides e.g., PRO826, PRO1068, PRO1184, PRO1346
PT and PRO1375, which stimulate proliferation of stimulated T-lymphocytes
PT are therapeutically useful for enhancing immune response and in cancer
PT treatments -
XX
PS Claim 2; Fig 152; 648pp; English.
XX
CC The invention describes an isolated human PRO polypeptide. The PRO
CC polypeptides are useful in detecting PRO polypeptides in a sample, in
CC linking a bioactive molecule to a cell expressing a PRO polypeptide, and
CC in modulating at least one biological activity of a cell expressing a PRO
CC polypeptide. PRO1312 stimulates hypertrophy of neonatal heart and is thus
CC useful for treating cardiac insufficiency disorders. PRO1154 and PRO1186
CC stimulate adrenal cortical capillary endothelial growth, and PRO536,
CC PRO943, PRO828, PRO826, PRO1068 or PRO535, PRO826, PRO819, PRO1126,
CC PRO1360 and PRO1387 induce c-fos in endothelial cells, and are thus
CC useful for treating conditions or disorders where angiogenesis would be
CC beneficial, e.g. wound healing and antagonist of this polypeptide are
CC useful for treating cancerous tumours. PRO812 inhibits vascular
CC endothelial growth factor (VEGF) stimulated proliferation of endothelial
CC cells and is thus useful for inhibiting endothelial cell growth in
CC mammals which would be beneficial in inhibiting tumour growth. PRO826,
CC PRO1068, PRO1184, PRO1346 and PRO1375 stimulate proliferation of
CC stimulated T-lymphocytes and are therapeutically useful for enhancing
CC immune response. PRO828, PRO826, PRO1068 or PRO1132 enhance survival of
CC retinal neurons cells (PRO1132 is also enhances survival/proliferation of
CC rod photoreceptor cells) and therefore are useful for treating retinal
CC disorders of injuries, e.g. retinitis pigmentosum, AMD. PRO819, PRO813
CC and PRO11066 induce proliferation of mammalian kidney mesangial cells,
CC and therefore are useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Berger disease or other
CC nephropathies associated with dermatitis, herpetiformis or Crohn's
CC disease. PRO1310, PRO844, PRO1312, PRO1192 and PRO1387 induce the
CC proliferation and/or redifferentiation of chondrocytes in culture and
CC are thus useful for treating sports injuries, and arthritis. This
CC sequence represents a novel human PRO protein polynucleotide.
XX
SQ Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
Db 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
Db 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
Db 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
Db 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
Db 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
Db 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy 361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
                    ||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 5
ABX80771
ID   ABX80771 standard; cDNA; 992 BP.
XX
AC   ABX80771;
XX
DT   22-APR-2003  (first entry)
XX
DE   Human secreted/transmembrane protein cDNA, #90.
XX
KW   Human; gene; ss; PRO; secreted; transmembrane; pharmaceutical;
KW   diagnostic; biosensor; bioreactor; tumour; therapeutic;
KW   gene therapy; tumour-associated antigenic target; TAT; ADEPT;
KW   antibody-dependent enzyme mediated prodrug therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003027162-A1.
XX
PD   06-FEB-2003.
XX
PF   15-NOV-2001; 2001US-0997428.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   30-MAR-2000; 2000WO-US08439.
PR   15-MAY-2000; 2000WO-US13358.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   11-AUG-2000; 2000WO-US22031.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   16-JUN-1997;   97US-049787P.
PR   17-OCT-1997;   97US-062250P.
PR   12-NOV-1997;   97US-065186P.
PR   13-NOV-1997;   97US-065311P.
PR   24-NOV-1997;   97US-066770P.
PR   25-FEB-1998;   98US-075945P.
PR   20-MAR-1998;   98US-078910P.
PR   28-APR-1998;   98US-083322P.
PR   07-MAY-1998;   98US-084600P.
PR   28-MAY-1998;   98US-087106P.
PR   02-JUN-1998;   98US-087607P.
PR   02-JUN-1998;   98US-087609P.
PR   02-JUN-1998;   98US-087759P.
PR   03-JUN-1998;   98US-087827P.
PR   04-JUN-1998;   98US-088021P.
PR   04-JUN-1998;   98US-088025P.
PR   04-JUN-1998;   98US-088026P.
PR   04-JUN-1998;   98US-088028P.
PR   04-JUN-1998;   98US-088029P.
PR   04-JUN-1998;   98US-088030P.
PR   04-JUN-1998;   98US-088033P.
PR   04-JUN-1998;   98US-088326P.
PR   05-JUN-1998;   98US-088167P.
PR   05-JUN-1998;   98US-088202P.
PR   05-JUN-1998;   98US-088212P.
PR   05-JUN-1998;   98US-088217P.
PR   09-JUN-1998;   98US-088655P.
PR   10-JUN-1998;   98US-088734P.
PR   10-JUN-1998;   98US-088738P.
PR   10-JUN-1998;   98US-088742P.
PR   10-JUN-1998;   98US-088810P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088826P.
PR   11-JUN-1998;   98US-088858P.
PR   11-JUN-1998;   98US-088861P.
PR   11-JUN-1998;   98US-088876P.
```

PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089947P.
PR 19-JUN-1998; 98US-089948P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 23-JUN-1998; 98US-090349P.
PR 23-JUN-1998; 98US-090355P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090431P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090445P.
PR 24-JUN-1998; 98US-090472P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 24-JUN-1998; 98US-090542P.
PR 24-JUN-1998; 98US-090557P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 01-JUL-1998; 98US-091360P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091519P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091633P.
PR 02-JUL-1998; 98US-091646P.
PR 02-JUL-1998; 98US-091673P.
PR 07-JUL-1998; 98US-091978P.
PR 07-JUL-1998; 98US-091982P.
PR 09-JUL-1998; 98US-092182P.
PR 10-JUL-1998; 98US-092472P.
PR 20-JUL-1998; 98US-093339P.
PR 30-JUL-1998; 98US-094651P.
PR 04-AUG-1998; 98US-095282P.
PR 04-AUG-1998; 98US-095285P.
PR 04-AUG-1998; 98US-095301P.
PR 04-AUG-1998; 98US-095302P.
PR 04-AUG-1998; 98US-095318P.
PR 04-AUG-1998; 98US-095321P.
PR 04-AUG-1998; 98US-095325P.
PR 10-AUG-1998; 98US-095916P.
PR 10-AUG-1998; 98US-095929P.
PR 10-AUG-1998; 98US-096012P.
PR 11-AUG-1998; 98US-096143P.
PR 11-AUG-1998; 98US-096146P.
PR 12-AUG-1998; 98US-096329P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096768P.
PR 17-AUG-1998; 98US-096773P.
PR 17-AUG-1998; 98US-096791P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096894P.
PR 17-AUG-1998; 98US-096895P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096950P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-096960P.
PR 18-AUG-1998; 98US-097022P.
PR 19-AUG-1998; 98US-097141P.
PR 20-AUG-1998; 98US-097218P.
PR 24-AUG-1998; 98US-097661P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-097978P.
PR 26-AUG-1998; 98US-097979P.
PR 26-AUG-1998; 98US-097986P.
PR 26-AUG-1998; 98US-098014P.
PR 31-AUG-1998; 98US-098525P.
PR 16-SEP-1998; 98US-100634P.
PR 17-SEP-1998; 98US-100858P.
PR 22-DEC-1998; 98US-113296P.
PR 12-MAR-1999; 99US-123957P.
PR 23-JUN-1999; 99US-141037P.

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
Db 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
Db 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
Db 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
Db 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
Db 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
Db 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy 361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
Db 361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy 421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
Db 421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy 481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
Db 481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy 541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
Db 541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy 601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
                ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 6
ABX81154
ID   ABX81154 standard; DNA; 992 BP.
XX
AC   ABX81154;
XX
DT   22-APR-2003  (first entry)
XX
DE   Novel human secreted or transmembrane protein PRO791 DNA.
XX
KW   Human; PRO; hypertrophy of neonatal heart; angiogenesis; wound healing;
KW   cardiac insufficiency disorder; cancer; tumour; immune response;
KW   adrenal cortical capillary endothelial growth; c-fos induction;
KW   vascular endothelial growth factor inhibition; VEGF inhibition;
KW   endothelial cell growth inhibitor; T-lymphocytes stimulation;
KW   retinal neurons cell survival; rod photoreceptor cell survival;
KW   retinal disorder; retinitis pigmentosum; kidney disorder;
KW   mammalian kidney mesangial cell proliferation; Berger disease;
KW   dermatitis; herpetiformis; Crohn's disease; chondrocyte proliferation;
KW   chondrocyte redifferentiation; sports injury; arthritis; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2003027985-A1.
XX
PD   06-FEB-2003.
XX
PF   14-NOV-2001; 2001US-0990562.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   30-MAR-2000; 2000WO-US08439.
PR   15-MAY-2000; 2000WO-US13358.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   28-JUL-2000; 2000WO-US20710.
PR   11-AUG-2000; 2000WO-US22031.
PR   23-AUG-2000; 2000WO-US23522.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   28-FEB-2001; 2001WO-US06520.
PR   01-JUN-2001; 2001WO-US17800.
PR   20-JUN-2001; 2001WO-US19692.
PR   29-JUN-2001; 2001WO-US21066.
PR   09-JUL-2001; 2001WO-US21735.
PR   16-JUN-1997;   97US-049787P.
PR   17-OCT-1997;   97US-062250P.
PR   12-NOV-1997;   97US-065186P.
PR   13-NOV-1997;   97US-065311P.
PR   24-NOV-1997;   97US-066770P.
PR   25-FEB-1998;   98US-075945P.
PR   20-MAR-1998;   98US-078910P.
PR   28-APR-1998;   98US-083322P.
PR   07-MAY-1998;   98US-084600P.
PR   28-MAY-1998;   98US-087106P.
PR   02-JUN-1998;   98US-087607P.
PR   02-JUN-1998;   98US-087609P.
PR   02-JUN-1998;   98US-087759P.
PR   03-JUN-1998;   98US-087827P.
PR   04-JUN-1998;   98US-088021P.
PR   04-JUN-1998;   98US-088025P.
PR   04-JUN-1998;   98US-088026P.
PR   04-JUN-1998;   98US-088028P.
PR   04-JUN-1998;   98US-088029P.
PR   04-JUN-1998;   98US-088030P.
PR   04-JUN-1998;   98US-088033P.
PR   04-JUN-1998;   98US-088326P.
PR   05-JUN-1998;   98US-088167P.
PR   05-JUN-1998;   98US-088202P.
PR   05-JUN-1998;   98US-088212P.
PR   05-JUN-1998;   98US-088217P.
PR   09-JUN-1998;   98US-088655P.
PR   10-JUN-1998;   98US-088734P.
PR   10-JUN-1998;   98US-088738P.
PR   10-JUN-1998;   98US-088742P.
PR   10-JUN-1998;   98US-088810P.
PR   10-JUN-1998;   98US-088824P.
PR   10-JUN-1998;   98US-088826P.
PR   11-JUN-1998;   98US-088858P.
PR   11-JUN-1998;   98US-088861P.
PR   11-JUN-1998;   98US-088876P.
PR   12-JUN-1998;   98US-089105P.
PR   16-JUN-1998;   98US-089440P.
PR   16-JUN-1998;   98US-089512P.
PR   16-JUN-1998;   98US-089514P.
PR   17-JUN-1998;   98US-089532P.
PR   17-JUN-1998;   98US-089538P.
PR   17-JUN-1998;   98US-089598P.
PR   17-JUN-1998;   98US-089599P.
PR   17-JUN-1998;   98US-089600P.
PR   17-JUN-1998;   98US-089653P.
```

PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089947P.
PR 19-JUN-1998; 98US-089948P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 23-JUN-1998; 98US-090349P.
PR 23-JUN-1998; 98US-090355P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090431P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090445P.
PR 24-JUN-1998; 98US-090472P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 24-JUN-1998; 98US-090542P.
PR 24-JUN-1998; 98US-090557P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 01-JUL-1998; 98US-091360P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091519P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091633P.
PR 02-JUL-1998; 98US-091646P.
PR 02-JUL-1998; 98US-091673P.
PR 07-JUL-1998; 98US-091978P.
PR 07-JUL-1998; 98US-091982P.
PR 09-JUL-1998; 98US-092182P.
PR 10-JUL-1998; 98US-092472P.
PR 20-JUL-1998; 98US-093339P.
PR 30-JUL-1998; 98US-094651P.
PR 04-AUG-1998; 98US-095282P.
PR 04-AUG-1998; 98US-095285P.
PR 04-AUG-1998; 98US-095301P.
PR 04-AUG-1998; 98US-095302P.
PR 04-AUG-1998; 98US-095318P.
PR 04-AUG-1998; 98US-095321P.
PR 04-AUG-1998; 98US-095325P.
PR 10-AUG-1998; 98US-095916P.
PR 10-AUG-1998; 98US-095929P.
PR 10-AUG-1998; 98US-096012P.
PR 11-AUG-1998; 98US-096143P.
PR 11-AUG-1998; 98US-096146P.
PR 12-AUG-1998; 98US-096329P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096768P.
PR 17-AUG-1998; 98US-096773P.
PR 17-AUG-1998; 98US-096791P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096894P.
PR 17-AUG-1998; 98US-096895P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096950P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-096960P.
PR 18-AUG-1998; 98US-097022P.
PR 19-AUG-1998; 98US-097141P.
PR 20-AUG-1998; 98US-097218P.
PR 24-AUG-1998; 98US-097661P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-097978P.
PR 26-AUG-1998; 98US-097979P.
PR 26-AUG-1998; 98US-097986P.
PR 26-AUG-1998; 98US-098014P.

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy   61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy  121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy  181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy  241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy  301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy  361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy  421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy  481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy  541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy  601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy  661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy  721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy  781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy  841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
                ||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 7
ABX90244
ID ABX90244 standard; cDNA; 992 BP.
XX
AC ABX90244;
XX
DT 01-MAY-2003 (first entry)
XX
DE Human secreted/transmembrane protein cDNA, #90.
XX
KW Human; gene; ss; PRO; secreted; transmembrane; signal peptide;
KW pharmaceutical; diagnostic; therapeutic; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002160384-A1.
XX
PD 31-OCT-2002.
XX
PF 14-NOV-2001; 2001US-0992598.
XX
PR 05-NOV-1997; 97WO-US20069.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 30-MAR-2000; 2000WO-US08439.
PR 15-MAY-2000; 2000WO-US13358.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.
PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 28-AUG-2001; 2001US-0941992.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI; 2003-288106/28.
DR P-PSDB; ABU60538.
XX
PT New transmembrane polypeptides and nucleic acids encoding the
PT polypeptides, useful in gene therapy, in chromosome identification, as
PT chromosome markers, or in generating probes -
XX
PS Claim 2; Fig 150; 650pp; English.
XX
CC The invention discloses isolated PRO secreted/transmembrane polypeptides
CC comprising a sequence without signal peptide and the nucleic acid

```
CC   encoding them. The polypeptides can be used to raise antibodies that
CC   specifically bind to the PRO polypeptide, for linking a bioactive
CC   molecule to a cell expressing a PRO protein and for modulating at least
CC   one biological activity of a cell. The PRO polypeptides or
CC   polynucleotides are also useful in gene therapy, in chromosome
CC   identification, as chromosome markers, or in generating probes. The PRO
CC   polypeptides are useful as molecular markers for protein
CC   electrophoresis, and the isolated nucleic acids may be used for
CC   recombinantly expressing those markers. The PRO polypeptides and nucleic
CC   acids may also be used in tissue typing. Anti-PRO antibodies are useful
CC   in diagnostic assays for PRO, and in affinity purification of PRO from
CC   recombinant cell culture or natural sources. The sequences presented in
CC   ABX90083-ABX90468 are the genes encoding, the primers amplifying and the
CC   probes detecting the PRO polynucleotides of the invention.
CC   Note: The sequence data for this patent is also available in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

  Query Match           100.0%;  Score 992;  DB 25;  Length 992;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-236;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 8
ABX77855
ID   ABX77855 standard; cDNA; 992 BP.
XX
AC   ABX77855;
XX
DT   14-APR-2003  (first entry)
XX
DE   Human PRO polynucleotide #61.
XX
KW   Human; PRO; gene; ss; cytostatic; tumour; cancer; breast; lung; stomach;
KW   liver; horse; cow; dog; cat; sheep; pig; goat; rabbit; ADEPT;
KW   antibody-dependent enzyme mediated prodrug therapy.
XX
OS   Homo sapiens.
XX
PN   US2003027163-A1.
XX
PD   06-FEB-2003.
XX
PF   15-NOV-2001; 2001US-0997666.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
PR   30-MAR-2000; 2000WO-US08439.
PR   15-MAY-2000; 2000WO-US13358.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
```

PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.
PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089947P.
PR 19-JUN-1998; 98US-089948P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 23-JUN-1998; 98US-090349P.
PR 23-JUN-1998; 98US-090355P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090431P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090445P.
PR 24-JUN-1998; 98US-090472P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 24-JUN-1998; 98US-090542P.
PR 24-JUN-1998; 98US-090557P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 01-JUL-1998; 98US-091360P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091519P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091633P.
PR 02-JUL-1998; 98US-091646P.
PR 02-JUL-1998; 98US-091673P.
PR 07-JUL-1998; 98US-091978P.
PR 07-JUL-1998; 98US-091982P.
PR 09-JUL-1998; 98US-092182P.
PR 10-JUL-1998; 98US-092472P.
PR 20-JUL-1998; 98US-093339P.
PR 30-JUL-1998; 98US-094651P.
PR 04-AUG-1998; 98US-095282P.
PR 04-AUG-1998; 98US-095285P.
PR 04-AUG-1998; 98US-095301P.
PR 04-AUG-1998; 98US-095302P.
PR 04-AUG-1998; 98US-095318P.
PR 04-AUG-1998; 98US-095321P.
PR 04-AUG-1998; 98US-095325P.
PR 10-AUG-1998; 98US-095916P.
PR 10-AUG-1998; 98US-095929P.
PR 10-AUG-1998; 98US-096012P.
PR 11-AUG-1998; 98US-096143P.
PR 11-AUG-1998; 98US-096146P.
PR 12-AUG-1998; 98US-096329P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096768P.
PR 17-AUG-1998; 98US-096773P.
PR 17-AUG-1998; 98US-096791P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096894P.
PR 17-AUG-1998; 98US-096895P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096950P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-096960P.
PR 18-AUG-1998; 98US-097022P.
PR 19-AUG-1998; 98US-097141P.
PR 20-AUG-1998; 98US-097218P.
PR 24-AUG-1998; 98US-097661P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-097978P.
PR 26-AUG-1998; 98US-097979P.
PR 26-AUG-1998; 98US-097986P.
PR 26-AUG-1998; 98US-098014P.
PR 31-AUG-1998; 98US-098525P.
PR 16-SEP-1998; 98US-100634P.
PR 17-SEP-1998; 98US-100858P.
PR 22-DEC-1998; 98US-113296P.
PR 12-MAR-1999; 99US-123957P.

PR 23-JUN-1999; 99US-141037P.
PR 07-JUL-1999; 99US-143048P.

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 9
ABX79451
ID ABX79451 standard; cDNA; 992 BP.
XX
AC ABX79451;
XX
DT 17-APR-2003 (first entry)
XX
DE Human sectreted/transmembrane protein cDNA, #90.
XX
KW Human; gene; ss; PRO; secreted; transmembrane; signal peptide;
KW pharmaceutical; diagnostic; biosensor; bioreactor; tumour; therapeutic;
KW colon cancer; lung cancer; breast cancer;cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002142961-A1.
XX
PD 03-OCT-2002.
XX
PF 19-NOV-2001; 2001US-0989721.
XX
PR 05-NOV-1997; 97WO-US20069.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 30-MAR-2000; 2000WO-US08439.
PR 15-MAY-2000; 2000WO-US13358.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.

PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 28-AUG-2001; 2001US-0941992.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI; 2003-155950/15.
DR P-PSDB; ABU58960.
XX
PT New secreted and transmembrane PRO polypeptides (e.g. PRO183, PRO184,
PT PRO361 or PRO846) useful as targets for therapeutic intervention in
PT cancers (e.g. lung or breast cancers), or for diagnosing these cancers
PT -
XX
PS Claim 2; Fig 150; 647pp; English.
XX
CC The invention discloses isolated PRO secreted/transmembrane polypeptides
CC comprising a sequence without signal peptide and the nucleic acid
CC encoding them. The polypeptides can be used to raise antibodies that
CC specifically bind to the PRO polypeptide, for linking a bioactive
CC molecule to a cell expressing a PRO protein and for modulating at least
CC one biological activity of a cell. The PRO polypeptides or
CC polynucleotides are also useful as pharmaceuticals, diagnostics,
CC biosensors or bioreactors, for detecting or treating e.g. tumours in
CC mammals, e.g. humans, dogs, cats, cattle, horses, sheep, pigs, goats or
CC rabbits as targets for therapeutic intervention in certain cancers (e.g.
CC colon, lung or breast cancers) and diagnostic determination of the
CC presence of these cancers. The PRO polypeptides are also useful as
CC molecular weight markers or for chromosome identification. The PRO genes
CC are useful as hybridisation probes or for screening libraries of human
CC cDNA, genomic DNA or mRNA. The PRO genes may also be used in gene
CC therapy, particularly for replacing a defective gene. The sequences
CC presented in ABX79290-ABX79675 are the genes encoding, the primers
CC amplifying and the probes detecting the PRO polynucleotides of the
CC invention.
CC Note: The sequence data for this patent is also available in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
```

```
Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 10
ABX64090
ID ABX64090 standard; cDNA; 992 BP.
XX
AC ABX64090;
XX
DT 26-FEB-2003 (first entry)
XX
DE cDNA encoding human PRO809 polypeptide.
XX
KW Human; PRO polypeptide; secreted protein; transmembrane protein;
KW genetic disorder; antibacterial; immunosuppressive; transgenic;
KW gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002103125-A1.
XX
PD 01-AUG-2002.
XX
PF 20-NOV-2001; 2001US-0989731.
XX
PR 05-NOV-1997; 97WO-US20069.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 05-JAN-1999; 99WO-US00106.
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 06-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000WO-US06319.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
PR 30-MAR-2000; 2000WO-US08439.
PR 15-MAY-2000; 2000WO-US13358.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.
PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 28-AUG-2001; 2001US-0941992.
XX
PA (GETH) GENENTECH LTD.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI; 2003-102117/09.
DR P-PSDB; ABU13920.
XX
PT Novel secreted and transmembrane polypeptide for modulating biological
PT activity of cell expressing the polypeptide, identifying agonists or
PT antagonists of polypeptide, and as molecular weight markers -
XX
PS Claim 2; Fig 150; 649pp; English.

```
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The
CC   PRO polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides are useful for detecting other PRO polypeptides, for
CC   linking bioactive molecules to cells expressing PRO polypeptides,
CC   for modulating biological activities of cells expressing PRO
CC   polypeptides, and for for identifying agonists or antagonists.
CC   The polynucleotide sequences encoding PRO polypeptides are useful as
CC   hybridisation probes, in chromosome and gene mapping, in the generation
CC   of antisense RNA and DNA, in the preparation of PRO polypeptides, for
CC   generating transgenic animals or knockout animals, to construct
CC   hybridisation probes for mapping the gene which encodes the PRO
CC   polypeptide, and for the genetic analysis of individuals with genetic
CC   disorders, in gene therapy, for chromosome identification, as
CC   chromosome markers, and for generating probes for PCR, Northern
CC   analysis, Southern analysis and Western analysis. The present
CC   sequence encodes a human PRO polypeptide of the invention.
CC   Note: The sequence data for this patent was obtained in electronic
CC   format directly from the USPTO web site at
CC   seqdata.uspto.gov/psipsDIDEntry.html.
XX
SQ   Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

  Query Match             100.0%;  Score 992;  DB 25;  Length 992;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-236;
  Matches  992;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992


RESULT 11
ABX17054
ID   ABX17054 standard; cDNA; 992 BP.
XX
AC   ABX17054;
XX
DT   04-FEB-2003  (first entry)
XX
DE   Human PRO polynucleotide #61.
XX
KW   Human; PRO; gene; ss; secreted polypeptide; transmembrane polypeptide;
KW   toxin; radiolabel; cell death; gene mapping; chromosome mapping;
KW   protein electrophoresis; genetic disorder; immunosuppressive; cytostatic;
KW   antibacterial.
XX
OS   Homo sapiens.
XX
PN   US2002123463-A1.
XX
PD   05-SEP-2002.
XX
PF   19-NOV-2001; 2001US-0989732.
XX
PR   05-NOV-1997;   97WO-US20069.
PR   16-SEP-1998;   98WO-US19330.
PR   17-SEP-1998;   98WO-US19437.
PR   07-OCT-1998;   98WO-US21141.
PR   01-DEC-1998;   98WO-US25108.
PR   05-JAN-1999;   99WO-US00106.
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   06-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   10-MAR-2000; 2000WO-US06319.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
```

PR 30-MAR-2000; 2000WO-US08439.
PR 15-MAY-2000; 2000WO-US13358.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 11-AUG-2000; 2000WO-US22031.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 16-JUN-1997; 97US-049787P.
PR 17-OCT-1997; 97US-062250P.
PR 12-NOV-1997; 97US-065186P.
PR 13-NOV-1997; 97US-065311P.
PR 24-NOV-1997; 97US-066770P.
PR 25-FEB-1998; 98US-075945P.
PR 20-MAR-1998; 98US-078910P.
PR 28-APR-1998; 98US-083322P.
PR 07-MAY-1998; 98US-084600P.
PR 28-MAY-1998; 98US-087106P.
PR 02-JUN-1998; 98US-087607P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088021P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088026P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088030P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088734P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088742P.
PR 10-JUN-1998; 98US-088810P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088858P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089440P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089532P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089599P.
PR 17-JUN-1998; 98US-089600P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089801P.
PR 18-JUN-1998; 98US-089907P.
PR 18-JUN-1998; 98US-089908P.
PR 28-AUG-2001; 2001US-0941992.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
DR WPI; 2003-066810/06.
DR P-PSDB; ABU10875.
XX
PT Novel secreted and transmembrane polypeptide for modulating biological
PT activity of cell expressing the polypeptide, identifying agonists or
PT antagonists of polypeptide, and as molecular weight markers -
XX
PS Claim 2; Fig 150; 655pp; English.
XX
CC The invention relates to a secreted and transmembrane polypeptide, termed
CC PRO polypeptide, and the polynucleotide encoding it. The polypeptide is
CC useful for detecting PRO polypeptides and for linking a bioactive
CC molecule to a cell expressing the above polypeptides, where the bioactive
CC molecule is a toxin, radiolabel or an antibody. The bioactive material
CC causes the death of the cell. The polypeptide is useful for identifying
CC agonists or antagonists of the PRO polypeptide, for preparing variants of
CC PRO, as a molecular weight marker for protein electrophoresis purposes
CC and the PRO polynucleotide is useful for recombinantly expressing those
CC markers. The polynucleotide is also useful as a hybridisation probe, in
CC chromosome and gene mapping, in generation of antisense RNA and DNA, in
CC the preparation of PRO polypeptide, for generating transgenic animals or
CC knockout animals which in turn are useful in the development and
CC screening of therapeutically useful reagents, to construct hybridisation
CC probes for mapping the gene which encodes PRO and for the genetic
CC analysis of individuals with genetic disorders, in gene therapy, for
CC chromosome identification, as a chromosome marker and for generating
CC probes for PCR, Northern analysis, Southern analysis and Western
CC analysis. This sequence represents a human PRO polynucleotide of the
CC invention.
XX
SQ Sequence 992 BP; 235 A; 301 C; 282 G; 174 T; 0 other;

Query Match 100.0%; Score 992; DB 25; Length 992;
Best Local Similarity 100.0%; Pred. No. 2.7e-236;
Matches 992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
Db 1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
Db 61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
Db 121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
Db 181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
Db 241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
Db 301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy 361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
Db 361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy 421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
Db 421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy 481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

```
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              ||||||||||||||||||||||||||||||||
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
```

RESULT 12
AAZ65030
ID AAZ65030 standard; cDNA; 991 BP.
XX
AC AAZ65030;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO809 encoding cDNA.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX
OS Homo sapiens.
XX
PN WO9963088-A2.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-US12252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.

```
PR   20-JUL-1998;   98US-0093339.
PR   30-JUL-1998;   98US-0094651.
PR   04-AUG-1998;   98US-0095282.
PR   04-AUG-1998;   98US-0095285.
PR   04-AUG-1998;   98US-0095301.
PR   04-AUG-1998;   98US-0095302.
PR   04-AUG-1998;   98US-0095318.
PR   04-AUG-1998;   98US-0095321.
PR   04-AUG-1998;   98US-0095325.
PR   10-AUG-1998;   98US-0095916.
PR   10-AUG-1998;   98US-0095929.
PR   10-AUG-1998;   98US-0096012.
PR   11-AUG-1998;   98US-0096143.
PR   11-AUG-1998;   98US-0096146.
PR   12-AUG-1998;   98US-0096329.
PR   17-AUG-1998;   98US-0096757.
PR   17-AUG-1998;   98US-0096766.
PR   17-AUG-1998;   98US-0096768.
PR   17-AUG-1998;   98US-0096773.
PR   17-AUG-1998;   98US-0096791.
PR   17-AUG-1998;   98US-0096867.
PR   17-AUG-1998;   98US-0096891.
PR   17-AUG-1998;   98US-0096894.
PR   17-AUG-1998;   98US-0096895.
PR   17-AUG-1998;   98US-0096897.
PR   18-AUG-1998;   98US-0096949.
PR   18-AUG-1998;   98US-0096950.
PR   18-AUG-1998;   98US-0096959.
PR   18-AUG-1998;   98US-0096960.
PR   18-AUG-1998;   98US-0097022.
PR   19-AUG-1998;   98US-0097141.
PR   20-AUG-1998;   98US-0097218.
PR   24-AUG-1998;   98US-0097661.
PR   26-AUG-1998;   98US-0097951.
PR   26-AUG-1998;   98US-0097952.
PR   26-AUG-1998;   98US-0097954.
PR   26-AUG-1998;   98US-0097955.
PR   26-AUG-1998;   98US-0097971.
PR   26-AUG-1998;   98US-0097974.
PR   26-AUG-1998;   98US-0097978.
PR   26-AUG-1998;   98US-0097979.
PR   26-AUG-1998;   98US-0097986.
PR   26-AUG-1998;   98US-0098014.
PR   31-AUG-1998;   98US-0098525.
PR   16-SEP-1998;   98US-0100634.
PR   12-JAN-1999;   99US-0115565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Chen J,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK;
PI   Wood WI,  Yuan J;
XX
DR   WPI; 2000-072883/06.
DR   P-PSDB; AAY66691.
XX
PT   Membrane-bound proteins and related nucleotide sequences  -
XX
PS   Claim 2; Fig 150; 822pp; English.
XX
CC   The invention provides membrane-bound PRO polypeptides and
CC   polynucleotides encoding them. The PRO sequences of the invention were
CC   identified based on extracellular domain homology screening. The PRO
CC   sequences have homology with proteins including LDL receptors, TIE
CC   ligands and various enzymes. The membrane-bound proteins and receptor
CC   molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC   immunoadhesins, for instance, can be used as therapeutic agents to block
CC   receptor-ligand interactions. The membrane-bound proteins can also be
CC   employed for screening of potential peptide or small molecule inhibitors
CC   of the relevant receptor/ligand interaction. The PRO encoding sequences
CC   are useful as hybridization probes, in chromosome and gene mapping and in
CC   the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC   will also be useful for the preparation of PRO polypeptides, especially
CC   by recombinant techniques.
XX
SQ   Sequence 991 BP; 234 A; 301 C; 282 G; 174 T; 0 other;

  Query Match             99.9%;  Score 991;  DB 21;  Length 991;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-236;
  Matches  991;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGCACGAGCCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAG 60

Qy         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGG 120

Qy        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCC 180

Qy        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCT 240

Qy        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGC 300

Qy        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCT 360

Qy        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGG 420

Qy        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGG 480

Qy        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACA 540

Qy        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG 600

Qy        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACA 660

Qy        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGACCAGAAGATGG 720

Qy        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCC 780

Qy        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGAC 840

Qy        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCA 900

Qy        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGC 960

Qy        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAAA 991
              ||||||||||||||||||||||||||||||
```

```
Db        961 TCTTGCCACAAAAAAAAAAAAAAAAAAAAA 991


RESULT 13
AAH64779
ID   AAH64779 standard; cDNA; 1047 BP.
XX
AC   AAH64779;
XX
DT   11-SEP-2001  (first entry)
XX
DE   Human secreted protein cDNA, SEQ ID NO: 55.
XX
KW   Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
KW   GENSET; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142451-A2.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-IB01938.
XX
PR   08-DEC-1999;   99US-0169629.
PR   06-MAR-2000; 2000US-0187470.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Bougueleret L,  Jobert S;
XX
DR   WPI; 2001-367870/38.
DR   P-PSDB; AAG89176.
XX
PT   Full length GENSET human nucleic acids encoding potentially secreted
PT   proteins, useful in gene therapy and vaccination against a variety of
PT   diseases, and for diagnosis of those diseases -
XX
PS   Claim 7; Page 620-621; 921pp; English.
XX
CC   The invention relates to full length GENSET human nucleic acids encoding
CC   potentially secreted proteins. The nucleic acids and the polypeptides
CC   they encode may be used in the prevention, treatment and diagnosis of
CC   diseases associated with inappropriate GENSET gene expression. For
CC   example, they be used to treat disorders associated with decreased
CC   GENSET gene expression by rectifying mutations or deletions in a
CC   patient's genome that affect the activity of GENSET or by supplementing
CC   the patients own production of GENSET polypeptides. Conversely,
CC   antisense nucleic acid molecules may be administered to down regulate
CC   GENSET expression by binding with the cells' own genes and preventing
CC   their expression. The sense and antisense nucleic acids may also be
CC   used as DNA probes in diagnostic assays to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples, and hence to
CC   determine which patients may be in need of restorative therapy.
CC   The GENSET polypeptides may be used as antigens in the production of
CC   antibodies and in assays to identify modulators (agonists and
CC   antagonists) of GENSET polypeptide expression and activity. The
CC   present sequence is a GENSET nucleic acid of the invention.
XX
SQ   Sequence 1047 BP; 235 A; 324 C; 302 G; 186 T; 0 other;

  Query Match           85.0%;  Score 843;  DB 22;  Length 1047;
  Best Local Similarity  90.7%;  Pred. No. 2.6e-199;
  Matches  950;  Conservative    0;  Mismatches    0;  Indels   97;  Gaps    1;

Qy         37 AGCAGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGC 96
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCAGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGC 60

Qy         97 TGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGGCTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTG 120

Qy        157 CCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCAC 216
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCAC 180

Qy        217 CCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCA 276
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCCAGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCA 240

Qy        277 AGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCA 336
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCA 300

Qy        337 GTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACA 396
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACA 360

Qy        397 GTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCA 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCA 420

Qy        457 ACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCT 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCT 480

Qy        517 CGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGC 576
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGGGCAGCCCACCTATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGC 540

Qy        577 AGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACT 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGAGACCATGCCACAGGCAGCCTGCCAACTTCTCCTTCCTGCCGAGCCAGACATCGGACT 600

Qy        637 GGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGC 696
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGTTCTGGTGCCAGGCTGCAAACAACGCCAATGTCCAGCACAGCGCCCTCACAGTGGTGC 660

Qy        697 CC---------------------------------------------------------- 698
              ||
Db        661 CCCCAGGAGGGTTGCCCAGGGCACCCACCATCGTGCTGGTTGGCAGCCTTGCCTCCACTG 720

Qy        699 ----------------------------------------CCAGGTGGTGACCAGAAGATG 719
                                                       |||||||||||||||||||||
Db        721 CGGCCATCACCTCCAGGATGCTGGGCTGGACCACGTGGGCCAGGTGGTGACCAGAAGATG 780

Qy        720 GAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACC 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAGGACTGGCAGGGTCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACC 840

Qy        780 CGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGA 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CGCCGTCTGAGTGAAGAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGA 900

Qy        840 CGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGC 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGC 960

Qy        900 AGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTG 959
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTG 1020

Qy        960 CTCTTGCCACAAAAAAAAAAAAAAAAA 986
              |||||||||||||||||||||||||||
Db       1021 CTCTTGCCACAAAAAAAAAAAAAAAAA 1047


RESULT 14
AAH99131
ID   AAH99131 standard; cDNA; 935 BP.
XX
AC   AAH99131;
XX
DT   12-OCT-2001  (first entry)
XX
```

```
DE   Human EST-derived coding sequence SEQ ID NO: 988.
XX
KW   Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW   tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW   diagnostics; forensic test; gene mapping; genetic disorder;
KW   biodiversity; gene therapy; nutrition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200154477-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02687.
XX
PR   25-JAN-2000; 2000US-0491404.
PR   17-JUL-2000; 2000US-0617746.
PR   03-AUG-2000; 2000US-0631451.
PR   15-SEP-2000; 2000US-0663870.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI   Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR   WPI; 2001-476164/51.
DR   P-PSDB; AAM24472.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Claim 1; Page 770; 1275pp; English.
XX
CC   The present invention provides the protein and coding sequences of novel
CC   proteins from a variety of organisms, including human, dog, cat, horse,
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a cDNA
CC   of the invention.
XX
SQ   Sequence 935 BP; 206 A; 286 C; 268 G; 175 T; 0 other;

  Query Match            79.7%;  Score 790.2;  DB 22;  Length 935;
  Best Local Similarity  94.6%;  Pred. No. 3.2e-186;
  Matches  920;  Conservative    0;  Mismatches    8;  Indels   45;  Gaps    8;

Qy          9 CCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAGGCATGGGG 68
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAGGAACTAGGAGGTTCTCACTGCCCGAGCAGAGGCCCTACACCCACCGAGGCATGGGG 60

Qy         69 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 120

Qy        129 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 180

Qy        189 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 240

Qy        249 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 300

Qy        309 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 360

Qy        369 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 420

Qy        429 TCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCC 488
              ||||                                        ||||||||||||||||||
Db        421 TCCA----------------------------------------GACAGAGGGGCAGGCCCC 442

Qy        489 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 502

Qy        549 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAAC-TT 607
              |||||||||||||||||||||||||||||||||||||||||||  ||||||||||| ||
Db        503 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGAGGCCTGCCAACTTT 562

Qy        608 CTCCTTCCTGCCGAGCCAGACATCGGAC-TGGTTCTGGTGCCAGGC-TGCAAACAACGCC 665
              |||||||||||||||||||||||||||| |||||||||||||||| |||||||||||||
Db        563 CTCCTTCCTGCCGAGCCAGACATCGGACTTGGTTCTGGTGCCAGGCTTGCAAACAACGCC 622

Qy        666 AATG-TCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGTGGTGA-CCAGAAGATGGAGG 723
              ||||  ||||||||||||||||||||||||  |||||||||||||| ||||||||||||||
Db        623 AATGTTCCAGCACAGCGCCCTCACAGTGGTTGCCCCAGGTGGTGACCCAGAAGATGGAGG 682

Qy        724 ACTGGCAGGGT-CCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGC 782
              |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        683 ACTGGCAGGGTCCCCCTGGAGAGCCCCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGC 742

Qy        783 CGTCTGAGTGAAGAGGAG-TTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACG 841
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        743 CGTCTGAGTGAAGAGGAGTTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACG 802

Qy        842 CAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAG 901
              |||||||||||||||||||  |||| ||||||||||||||||||||||||||||| |||
Db        803 CAAAGCAGCAGCCATGTAGATTGAATCGTCCAGAGAGCCAAGCACGGCAGAGGACTTCAG 862

Qy        902 GCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCT 961
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        863 GCCATCAGCGTGCACTGTTCGTATTTGGGGTTCATGCAAAATGAGTGTGTTTTAGCTGCT 922

Qy        962 CTTGCCACAAAAA 974
              |||||||||||||
Db        923 CTTGCCACAAAAA 935
```

```
RESULT 15
AAC74260
ID   AAC74260 standard; cDNA; 1047 BP.
XX
AC   AAC74260;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Human secreted protein gene 38 SEQ ID NO:48.
XX
KW   Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW   antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW   nootropic; neuroprotective; antibacterial; virucide; fungicide; neoplasm;
KW   opthalmalogical; autoimmune disease; rheumatoid arthritis; angiogenesis;
KW   hyperproliferative disorder; cardiovascular disorder; infection;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; chemotaxis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200056754-A1.
XX
PD   28-SEP-2000.
XX
PF   16-MAR-2000; 2000WO-US06792.
XX
PR   19-MAR-1999;   99US-0125362.
PR   10-DEC-1999;   99US-0169980.
XX
```

```
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen GA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-579483/54.
DR   P-PSDB; AAB39216.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition -
XX
PS   Claim 1; Page 363-364; 434pp; English.
XX
CC   The polynucleotide sequences given in AAC74223-C74279 encode the human
CC   secreted proteins represented in AAB39179-B39226. Sequences
CC   AAB39227-B39308 are alternative proteins encoded by the genes, and also
CC   protein sequences with which they share homology. The proteins have
CC   activities based on the tissues and cells in which they are expressed.
CC   Examples of activities include: immunosuppressive; antiarthritic;
CC   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
CC   fungicide; and opthalmalogical. The human secreted proteins,
CC   polynucleotides, antagonists and agonists of the invention may be useful
CC   in the treatment, prevention, and/or diagnosis of various disease,
CC   disorders and conditions such as autoimmune diseases e.g. rheumatoid
CC   arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC   liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC   disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC   e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC   fungi and ocular disorders e.g. corneal infection. The polypeptides can
CC   also be used to aid wound healing and epithelial cell proliferation, to
CC   regenerate tissues, maintain organs before transplantation, in
CC   chemotaxis and as a food additive or preservative e.g. to increase
CC   storage capabilities. Sequences AAC74214-C74222 and AAB39178 are used
CC   during the isolation and characterisation of the genes of the invention.
XX
SQ   Sequence 1047 BP; 238 A; 323 C; 299 G; 184 T; 3 other;

  Query Match             78.0%;  Score 773.6;  DB 21;  Length 1047;
  Best Local Similarity   87.1%;  Pred. No. 4.4e-182;
  Matches  941;  Conservative    1;  Mismatches    0;  Indels  138;  Gaps    3;

Qy         10 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGA-GGCCCTACACCCACCGAGGCATGGGG 68
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGANGGCCCTACACCCACCGAGGCATGGGG 60

Qy         69 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 120

Qy        129 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 180

Qy        189 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 240

Qy        249 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 300

Qy        309 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 360

Qy        369 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 420

Qy        429 TCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCC 488
              ||||                                        ||||||||||||||||||
Db        421 TCCA----------------------------------------GACAGAGGGGCAGGCCCC 442

Qy        489 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 502

Qy        549 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 562

Qy        609 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 622

Qy        669 GTCCAGCACAGCGCCCTCACAGTGGTGCC---------------------------------- 697
              ||||||||||||||||||||||||||||
Db        623 GTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGAGGGTTSCCCAGGGCACCCACCATC 682

Qy        698 ---------------------------------------------------------- 697

Db        683 GTGCTGGTTGGCAGCCTTGCCTCCACTGCGGCCATCACCTCCAGGATGCTGGGCTGGACC 742

Qy        698 --------CCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 749
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 CACGTGGGCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 802

Qy        750 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 862

Qy        810 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        863 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACYG 922

Qy        870 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 982

Qy        930 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 1042

RESULT 16
ABT16794
ID   ABT16794 standard; DNA; 1047 BP.
XX
AC   ABT16794;
XX
DT   03-APR-2003  (first entry)
XX
DE   Human secreted protein gene sequence - SEQ ID No 43.
XX
KW   Human; gene; ds; protein therapy; immediate hypersensitivity disease;
KW   allergic disorder; asthmatic disorder; gene therapy; secreted protein;
KW   hay fever; allergic conjunctivitis; allergic rhinitis;
KW   binding partner identification; chromosome identification;
KW   radiation hybrid mapping; long-range restriction mapping.
XX
OS   Homo sapiens.
XX
PN   WO200277188-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US09239.
XX
PR   27-MAR-2001; 2001US-278650P.
PR   12-SEP-2001; 2001US-0950082.
PR   12-SEP-2001; 2001US-0950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
```

```
DR   WPI; 2003-175010/17.
XX
PT   Use of human secreted proteins and nucleic acids for preparing a
PT   diagnostic or pharmaceutical composition for diagnosing or treating
PT   allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic
PT   conjunctivitis or rhinitis -
XX
PS   Claim 7; Page 586; 823pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of human
CC   secreted proteins. The DNA and protein sequences of the invention are
CC   useful for the diagnosis and treatment of allergic disorders, asthmatic
CC   disorders and immediate hypersensitivity diseases (e.g. hay fever,
CC   allergic conjunctivitis and allergic rhinitis). The proteins of the
CC   invention are also useful for identifying a binding partner. The nucleic
CC   acids of the invention are also useful for chromosome identification,
CC   radiation hybrid mapping or long-range restriction mapping. The present
CC   DNA sequence encodes a human secreted protein of the invention.
XX
SQ   Sequence 1047 BP; 238 A; 323 C; 299 G; 184 T; 3 other;

  Query Match             78.0%;  Score 773.6;  DB 25;  Length 1047;
  Best Local Similarity   87.1%;  Pred. No. 4.4e-182;
  Matches  941;  Conservative    1;  Mismatches    0;  Indels  138;  Gaps    3;

Qy         10 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGA-GGCCCTACACCCACCGAGGCATGGGG 68
              ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db          1 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGANGGCCCTACACCCACCGAGGCATGGGG 60

Qy         69 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 120

Qy        129 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 180

Qy        189 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 240

Qy        249 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 300

Qy        309 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 360

Qy        369 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 420

Qy        429 TCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCC 488
              ||||                                        ||||||||||||||||||
Db        421 TCCA----------------------------------------GACAGAGGGGCAGGCCCC 442

Qy        489 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 502

Qy        549 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 562

Qy        609 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 622

Qy        669 GTCCAGCACAGCGCCCTCACAGTGGTGCC---------------------------------- 697
              |||||||||||||||||||||||||||
Db        623 GTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGAGGGTTSCCCAGGGCACCCACCATC 682

Qy        698 ---------------------------------------------------------- 697

Db        683 GTGCTGGTTGGCAGCCTTGCCTCCACTGCGGCCATCACCTCCAGGATGCTGGGCTGGACC 742

Qy        698 --------CCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 749
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 CACGTGGGCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 802

Qy        750 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 862

Qy        810 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCG 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        863 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACYG 922

Qy        870 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 982

Qy        930 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 1042
```

```
RESULT 17
ABZ66993
ID   ABZ66993 standard; cDNA; 1047 BP.
XX
AC   ABZ66993;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Human secreted protein encoding cDNA SEQ ID NO 113.
XX
KW   Human; secreted protein; nootropic; neuroprotective; cytostatic;
KW   virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   vulnerary; antibacterial; antiparkinsonian; antisickling; antianaemic;
KW   antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW   antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;
KW   antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW   cardiovascular disorder; neurological disease; nephrotropic;
KW   gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277186-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-US09188.
XX
PR   27-MAR-2001; 2001US-278650P.
PR   12-SEP-2001; 2001US-0950082.
PR   12-SEP-2001; 2001US-0950083.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2003-040583/03.
DR   P-PSDB; ABP99572.
XX
PT   New human secreted proteins encoded by genes contained in cDNA clones
PT   (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g.
PT   AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne
PT   encephalitis or West Nile fever -
XX
PS   Claim 7; Page 1235-1236; 2423pp; English.
XX
CC   The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC   encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC   treating or ameliorating medical conditions e.g. by protein or gene
```

```
CC   therapy. The genes are isolated from a range of human tissues disclosed
CC   in the specification. The nucleic acids, proteins, antibodies and
CC   (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC   (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC   adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC   lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC   allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC   diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC   arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC   myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC   cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC   bacterial, fungal and parasitic infections.
XX
SQ   Sequence 1047 BP; 238 A; 323 C; 299 G; 184 T; 3 other;

  Query Match             78.0%;  Score 773.6;  DB 25;  Length 1047;
  Best Local Similarity   87.1%;  Pred. No. 4.4e-182;
  Matches  941;  Conservative    1;  Mismatches    0;  Indels  138;  Gaps    3;

Qy         10 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGA-GGCCCTACACCCACCGAGGCATGGGG 68
               |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db          1 CAGGAACTAGGAGGTTCTCACTGCCCGAGCAGANGGCCCTACACCCACCGAGGCATGGGG 60

Qy         69 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 128
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGCTTCTCCAAGGCACGGGAG 120

Qy        129 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 188
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGC 180

Qy        189 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 248
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTC 240

Qy        249 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 308
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGTGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCC 300

Qy        309 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 368
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCG 360

Qy        369 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 428
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGG 420

Qy        429 TCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCC 488
               ||||                                  ||||||||||||||||||
Db        421 TCCA------------------------------------GACAGAGGGGCAGGCCCC 442

Qy        489 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 548
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATC 502

Qy        549 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 608
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 GGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTC 562

Qy        609 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 668
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TCCTTCCTGCCGAGCCAGACATCGGACTGGTTCTGGTGCCAGGCTGCAAACAACGCCAAT 622

Qy        669 GTCCAGCACAGCGCCCTCACAGTGGTGCC---------------------------- 697
               |||||||||||||||||||||||||||
Db        623 GTCCAGCACAGCGCCCTCACAGTGGTGCCCCCAGGAGGGTTSCCCAGGGCACCCACCATC 682

Qy        698 ------------------------------------------------------------ 697

Db        683 GTGCTGGTTGGCAGCCTTGCCTCCACTGCGGCCATCACCTCCAGGATGCTGGGCTGGACC 742

Qy        698 --------CCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 749
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        743 CACGTGGGCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCCCC 802

Qy        750 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 809
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGGGG 862

Qy        810 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCG 869
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        863 TTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACYG 922

Qy        870 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 929
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGG 982

Qy        930 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 989
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAA 1042
```

```
RESULT 18
ABZ36822
ID   ABZ36822 standard; cDNA; 502 BP.
XX
AC   ABZ36822;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Human GENSET coding sequence SEQ ID 695.
XX
KW   Cytostatic; antiinflammatory; nootropic; neuroprotective; cardiant;
KW   gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;
KW   inflammatory disease; immune disorder; neuromuscular; toxicity;
KW   central nervous system; cardiovascular; gastrointestinal; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200283898-A1.
XX
PD   24-OCT-2002.
XX
PF   18-APR-2001; 2001WO-IB00914.
XX
PR   18-APR-2001; 2001WO-IB00914.
XX
PA   (GEST ) GENSET.
XX
PI   Bejanin S,  Tanaka H,  Dumas Milne Edwards J,  Jobert S,  Giordano J;
XX
DR   WPI; 2003-075548/07.
XX
PT   New GENSET polynucleotides and polypeptides, useful for treating heavy
PT   metal toxicity, cancer, inflammatory diseases, immune disorders, and
PT   the neuromuscular, CNS, cardiovascular or gastrointestinal effects of
PT   the toxicity  -
XX
PS   Claim 12; Page 656; 735pp; English.
XX
CC   The present invention relates to novel GENSET polynucleotides
CC   (ABZ36404-ABZ36911) encoding polypeptides (ABP75963-ABP76368). The
CC   polynucleotides and polypeptides are useful in screening and diagnostic
CC   assays for abnormal GENSET expression and/or biological activity. They
CC   are also useful for screening of compounds for treating or preventing
CC   GENSET-related disorders, such as heavy metal toxicity, cancer,
CC   inflammatory diseases, immune disorders, and the neuromuscular, central
CC   nervous system (CNS), cardiovascular or gastrointestinal effects of the
CC   toxicity. The polynucleotides are useful for constructing or expanding
CC   chromosome maps.
XX
SQ   Sequence 502 BP; 105 A; 169 C; 135 G; 92 T; 1 other;

  Query Match             50.6%;  Score 501.6;  DB 25;  Length 502;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-114;
  Matches  501;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy         40 AGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGG 99
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAGGCCCTACACCCACCGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGG 60

Qy        100 CTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCT 159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCCAGCAGCTTCTCCAAGGCACGGGAGGAAGAAATTACCCCTGTGGTCTCCATTGCCT 120

Qy        160 ACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCC 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACAAAGTCCTGGAAGTTTTCCCCAAAGGCCGCTGGGTGCTCATAACCTGCTGTGCACCCC 180

Qy        220 AGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGA 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCACCACCGCCCATCACCTATTCCCTCTGTGGAACCAAGAACATCAAGGTGGCCAAGA 240

Qy        280 AGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTC 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACCTCAACGTCACACTCAAGTCCAGTC 300

Qy        340 CAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTG 399
              ||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db        301 CAGACCTGCTCACCTACTTCTGCYGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAGTG 360

Qy        400 CCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACT 459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACT 420

Qy        460 TCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGG 519
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGG 480

Qy        520 GCAGCCCACCTATCACCAACAG 541
              |||||||||||||||||||||
Db        481 GCAGCCCACCTATCACCAACAG 502
```

RESULT 19
AAS34076
ID AAS34076 standard; cDNA; 724 BP.
XX
AC AAS34076;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cDNA encoding a novel foetal antigen, SEQ ID No 600.
XX
KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200155312-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01321.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488782/53.
DR P-PSDB; AAU21256.
XX
PT New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems -
XX
PS Claim 1; SEQ ID No 600; 642pp; English.
XX
CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 50.6%; Score 501.6; DB 22; Length 724;
Best Local Similarity 98.0%; Pred. No. 1.2e-114;
Matches 545; Conservative 4; Mismatches 3; Indels 4; Gaps 4;

```
Qy        131 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 190
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 AGAAATTACCCCTGTGGTCTCCATTGCCTACAAAGTCCTGGAAGTTTTCCCCAAAGGCCG 220

Qy        191 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 250
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTGGGTGCTCATAACCTGCTGTGCACCCCAGCCACCACCGCCCATCACCTATTCCCTCTG 280

Qy        251 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 TGGAACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTT 340

Qy        311 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTC 370
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CAACCTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTC 400

Qy        371 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 CTCCACCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTC 460

Qy        431 CAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAG 490
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        461 CAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGG-CCCAG 519

Qy        491 GGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACCTATCACCAACAGCCTGATCGG 550
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        520 GGTGGAGATGATCTGCCAGGCGTCCTCGGGCAG-CCACCTATCACCAACAGCCTGATCGG 578

Qy        551 GAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTGCCAACTTCTC 610
              ||||||||||||||:|||||||||||||||||||||||||||||||| ||||||:||
Db        579 GAAGGATGGGCAGGKCCACCTGCAGCAGAGACCATGCCACAGGCAGCCTG-CAACTTSTC 637

Qy        611 CTTCCTGCCGAGCCAGACATCGGACTGGTTC-TGGTGCCAGGCTGCAAACAACGCCAATG 669
              |||||||||||||||||||||||||||||: |||||||||| ||||||||||:|||||
Db        638 CTTCCTGCCGAGCCAGACATCGGACTGGTTYTTGGTGCCAGGGTGCAAACAACGYCAATG 697

Qy        670 TCCAGCACAGCGCCCT 685
              |||||||||| | |||
Db        698 TCCAGCACAGGGGCCT 713
```

RESULT 20
ABA20618
ID ABA20618 standard; DNA; 32220 BP.
XX
AC ABA20618;
XX

DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12949.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 12949; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32220 BP; 7879 A; 8016 C; 7968 G; 8357 T; 0 other;

Query Match 29.5%; Score 292.8; DB 22; Length 32220;
Best Local Similarity 97.7%; Pred. No. 2.8e-62;
Matches 297; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       688 CAGTGGTGCCCCCAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 747
            || || |   | ||||||||||||||||||||||||||||||||||||||||||||||||
Db      9199 CACTGCTTTGTCAAGGTGGTGACCAGAAGATGGAGGACTGGCAGGGTCCCCTGGAGAGCC 9258

Qy       748 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 807
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9259 CCATCCTTGCCTTGCCGCTCTACAGGAGCACCCGCCGTCTGAGTGAAGAGGAGTTTGGGG 9318

Qy       808 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 867
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9319 GGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAC 9378

Qy       868 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 927
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9379 CGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTT 9438

Qy       928 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAA 987
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9439 GGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAA 9498

Qy       988 AAAA 991
            ||||
Db      9499 AAAA 9502
```

RESULT 21
AAX12008/c
ID AAX12008 standard; DNA; 150 BP.
XX
AC AAX12008;
XX
DT 30-MAR-1999 (first entry)
XX
DE Human biallelic polymorphic DNA fragment EST376246b.
XX
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
XX
OS Homo sapiens.
XX
PN WO9820165-A2.
XX
PD 14-MAY-1998.
XX
PF 05-NOV-1997; 97WO-US20313.
XX
PR 06-NOV-1996; 96US-0030455.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Hudson T, Lander ES, Wang D;
XX
DR WPI; 1998-286974/25.
XX
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
XX
PS Claim 1; Page 212; 310pp; English.
XX
CC AAX10269-X12937 are human DNA fragments which contain biallelic
CC polymorphic markers which have been isolated using the primers
CC represented in AAX09121-X10268. The base occupying the polymorphic site
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC can be used in methods for determining polymorphic forms in an individual
CC for use in e.g. forensics, paternity testing or for phenotypic typing for
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC familial hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
XX
SQ Sequence 150 BP; 30 A; 47 C; 32 G; 40 T; 1 other;

Query Match 15.1%; Score 149.6; DB 19; Length 150;
Best Local Similarity 99.3%; Pred. No. 1.6e-27;
Matches 149; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
            ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db       150 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAAYCGTCCAGAGAGC 91

Qy       880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        90 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 31

Qy       940 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
            ||||||||||||||||||||||||||||||
```

```
Db         30 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1
```

RESULT 22
AAX12009/c
ID AAX12009 standard; DNA; 150 BP.
XX
AC AAX12009;
XX
DT 30-MAR-1999 (first entry)
XX
DE Human biallelic polymorphic DNA fragment EST376246a.
XX
KW Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW detection; phenotypic typing; characteristic; infection; hereditary;
KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW treatment; marker; ss.
XX
OS Homo sapiens.
XX
PN WO9820165-A2.
XX
PD 14-MAY-1998.
XX
PF 05-NOV-1997; 97WO-US20313.
XX
PR 06-NOV-1996; 96US-0030455.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Hudson T, Lander ES, Wang D;
XX
DR WPI; 1998-286974/25.
XX
PT New isolated nucleic acid segments from the human genome - used for
PT determining polymorphic forms for use in e.g. forensics, paternity
PT testing or phenotypic typing for disease
XX
PS Claim 1; Page 212; 310pp; English.
XX
CC AAX10269-X12937 are human DNA fragments which contain biallelic
CC polymorphic markers which have been isolated using the primers
CC represented in AAX09121-X10268. The base occupying the polymorphic site
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC can be used in methods for determining polymorphic forms in an individual
CC for use in e.g. forensics, paternity testing or for phenotypic typing for
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC familial hypercholesterolemia, polycystic kidney disease, hereditary
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC autoimmune diseases, inflammation, cancer, diseases of the nervous
CC system, infection by pathogenic microorganisms, and characteristics such
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC endurance, fertility, and susceptibility or receptivity to particular
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC segments can also be used to produce medicaments for the treatment or
CC prophylaxis of such diseases.
XX
SQ Sequence 150 BP; 30 A; 46 C; 33 G; 40 T; 1 other;

Query Match 15.1%; Score 149.6; DB 19; Length 150;
Best Local Similarity 99.3%; Pred. No. 1.6e-27;
Matches 149; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy        820 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 879
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 GGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAGC 91

Qy        880 CAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCA 939
              |||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db         90 CAAGCACGGCAGAGGACTGCAGGCCATCAGCRTGCACTGTTCGTATTTGGAGTTCATGCA 31
```

```
Qy        940 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 969
              ||||||||||||||||||||||||||||||
Db         30 AAATGAGTGTGTTTTAGCTGCTCTTGCCAC 1
```

RESULT 23
ABX97141
ID ABX97141 standard; cDNA; 1272 BP.
XX
AC ABX97141;
XX
DT 20-MAY-2003 (first entry)
XX
DE Human NOV93a cDNA.
XX
KW NOVX; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
KW hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200272757-A2.
XX
PD 19-SEP-2002.
XX
PF 08-MAR-2002; 2002WO-US06908.
XX
PR 08-MAR-2001; 2001US-274101P.
PR 08-MAR-2001; 2001US-274194P.
PR 08-MAR-2001; 2001US-274281P.
PR 08-MAR-2001; 2001US-274322P.
PR 09-MAR-2001; 2001US-274849P.
PR 12-MAR-2001; 2001US-275235P.
PR 13-MAR-2001; 2001US-275578P.
PR 13-MAR-2001; 2001US-275579P.
PR 13-MAR-2001; 2001US-275601P.
PR 14-MAR-2001; 2001US-276000P.
PR 16-MAR-2001; 2001US-276776P.
PR 19-MAR-2001; 2001US-276994P.
PR 20-MAR-2001; 2001US-277239P.
PR 20-MAR-2001; 2001US-277321P.
PR 20-MAR-2001; 2001US-277327P.
PR 21-MAR-2001; 2001US-277791P.
PR 22-MAR-2001; 2001US-277833P.
PR 23-MAR-2001; 2001US-278152P.
PR 26-MAR-2001; 2001US-278894P.
PR 27-MAR-2001; 2001US-278999P.
PR 27-MAR-2001; 2001US-279036P.
PR 28-MAR-2001; 2001US-279344P.
PR 30-MAR-2001; 2001US-277338P.
PR 30-MAR-2001; 2001US-279995P.
PR 30-MAR-2001; 2001US-280233P.
PR 02-APR-2001; 2001US-280802P.
PR 02-APR-2001; 2001US-280822P.
PR 02-APR-2001; 2001US-280900P.
PR 04-APR-2001; 2001US-281194P.
PR 13-APR-2001; 2001US-283675P.
PR 30-APR-2001; 2001US-287424P.
PR 02-MAY-2001; 2001US-288066P.
PR 03-MAY-2001; 2001US-288342P.
PR 03-MAY-2001; 2001US-288528P.
PR 15-MAY-2001; 2001US-291190P.
PR 16-MAY-2001; 2001US-291099P.
PR 16-MAY-2001; 2001US-291240P.
PR 30-MAY-2001; 2001US-294485P.
PR 31-MAY-2001; 2001US-294889P.
PR 31-MAY-2001; 2001US-294899P.
PR 18-JUN-2001; 2001US-299027P.
PR 19-JUN-2001; 2001US-299303P.
PR 19-JUN-2001; 2001US-299310P.
PR 10-JUL-2001; 2001US-304354P.
PR 31-JUL-2001; 2001US-309198P.

```
PR   16-AUG-2001; 2001US-312903P.
PR   10-SEP-2001; 2001US-318462P.
PR   12-SEP-2001; 2001US-318770P.
PR   27-SEP-2001; 2001US-325430P.
PR   27-SEP-2001; 2001US-325681P.
PR   18-OCT-2001; 2001US-330380P.
PR   31-OCT-2001; 2001US-335301P.
PR   14-NOV-2001; 2001US-332172P.
PR   14-NOV-2001; 2001US-332271P.
PR   14-NOV-2001; 2001US-332272P.
PR   14-NOV-2001; 2001US-333184P.
PR   14-NOV-2001; 2001US-333272P.
PR   21-NOV-2001; 2001US-332094P.
PR   03-DEC-2001; 2001US-337426P.
PR   03-DEC-2001; 2001US-338092P.
PR   04-DEC-2001; 2001US-337185P.
PR   03-JAN-2002; 2002US-345705P.
PR   07-MAR-2002; 2002US-0092900.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Padigaru M,  Spytek KA,  Shenoy SG,  Taupier RJ,  Pena CEA,  Li L;
PI   Zerhusen BD,  Gusev V,  Ji W,  Gorman L,  Miller CE,  Kekuda R;
PI   Patturajan M,  Gangolli E,  Vernet CAM,  Guo X,  Tchernev V;
PI   Fernandes ER,  Casman SJ,  Malyankar UM,  Gerlach V,  Liu Y;
PI   Anderson D,  Spaderna SK,  Catterton E,  Burgess C,  Leite M,  Zhong H;
PI   Alsobrook JP,  Lepley DM,  Rieger DK;
XX
DR   WPI; 2002-723332/78.
DR   P-PSDB; ABU65174.
XX
PT   NOVX polypeptides and polynucleotides, useful for preventing or
PT   treating a disorder associated with aberrant NOVX expression or
PT   activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or
PT   bronchial asthma
XX
PS   Claim 13; Page 368-369; 1103pp; English.
XX
CC   This invention describes novel human NOVX polypeptides which have
CC   cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and
CC   hypotensive activity. Pharmaceutical compositions comprising the NOVX
CC   proteins or nucleic acid molecules or NOVX antibodies are useful for
CC   preventing or treating a disorder associated with aberrant NOVX
CC   expression or activity e.g. cancer, hypertension, atherosclerosis,
CC   cardiomyopathy or bronchial asthma. The products of the invention can
CC   be used for gene therapy or in a vaccine. ABX97008-ABX97185 are cDNA
CC   fragments amplified and isolated by the PCR primers and probes
CC   represented in ABX13460-ABX13462 and ABX97186-ABX97593. ABX97008-ABX97185
CC   encode the NOVX proteins described in ABU65041-ABU65218.
XX
SQ   Sequence 1272 BP; 182 A; 521 C; 356 G; 213 T; 0 other;

  Query Match             4.8%;  Score 47.2;  DB 24;  Length 1272;
  Best Local Similarity   48.8%;  Pred. No. 0.084;
  Matches  127;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy        256 CCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAACC 315
              ||  | |||  ||| ||  |  |  ||||  |||  ||       | ||||  || | ||
Db        711 CCCCGGACACGAAGCTGATCGTGGTGGTGCGGAACCCCGTGACCCGGGCCATCTCCGACT 770

Qy        316 TCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCA 375
                  |   || |||       |  || | ||||| || || |  |  || ||| | | |
Db        771 AGGCCCAGACGCTCTCCAAGACCCCGGGCCTGCCCAGCTTCCGCGCCCTGGCCTTCCGCC 830

Qy        376 CCTCAGGTGCCCATGTGGACAGTGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGC 435
                |      | ||  |||||||  ||| ||       | | |  |  ||||  || |||
Db        831 ACGGCCTGGGCCCCGTGGACACAGCCTGGAGCGCCGTCCGCATCGGCCTGTACGCCCAGC 890

Qy        436 CAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGG 495
              ||    |  || |  | ||||| | |||    ||      |  | |  | |||| |
Db        891 ACCTGGACCACTGGCTGCGCTACTTCCCCCTGTCCCACTTCCTGTTCGTCAGCGGGGAGC 950

Qy        496 AGATGATCTGCCAGGCGTCC 515
```

```
                ||  ||  ||  |  ||  ||
Db        951 GTCTGGTCAGCGACCCGGCC 970


RESULT 24
ABQ76896
ID   ABQ76896 standard; DNA; 53226 BP.
XX
AC   ABQ76896;
XX
DT   13-MAR-2003  (first entry)
XX
DE   Human G-protein coupled receptor DNA SEQ ID 3.
XX
KW   G-protein coupled receptor; secretin receptor subfamily; human; SNP;
KW   GPCR; protease; Parkinson's disease; gene; chromosome X;
KW   single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3000..50651
FT                   /*tag=  a
FT                   /product= "GPCR"
FT                   /note "this coding sequence is interrupted by
FT                         13 introns"
FT   variation       replace (1746,c)
FT                   /*tag=  b
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (1755,g)
FT                   /*tag=  c
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (1961,g)
FT                   /*tag=  d
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            3000..3088
FT                   /*tag=  e
FT                   /number= 1
FT   intron          3089..3874
FT                   /*tag=  f
FT                   /number= 1
FT   exon            3875..4038
FT                   /*tag=  g
FT                   /number= 2
FT   intron          4039..6037
FT                   /*tag=  h
FT                   /number= 2
FT   variation       replace (5411,g)
FT                   /*tag=  i
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (5760,a)
FT                   /*tag=  j
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            6038..6170
FT                   /*tag=  k
FT                   /number= 3
FT   intron          6171..8059
FT                   /*tag=  l
FT                   /number= 3
FT   exon            8060..8178
FT                   /*tag=  m
FT                   /number= 4
FT   intron          8179..15910
FT                   /*tag=  n
FT                   /number= 4
FT   variation       replace (11390,c)
FT                   /*tag=  o
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (11390,t)
FT                   /*tag=  p
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            15911..16127
FT                   /*tag=  q
```

```
FT                   /number= 5
FT   intron          16128..17484
FT                   /*tag=  r
FT                   /number= 5
FT   variation       replace (16988,a)
FT                   /*tag=  s
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            17485..17647
FT                   /*tag=  t
FT                   /number= 6
FT   intron          17648..32332
FT                   /*tag=  u
FT                   /number= 6
FT   variation       replace (18361,g)
FT                   /*tag=  v
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (19769,c)
FT                   /*tag=  w
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (22910,t)
FT                   /*tag=  x
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (22935,t)
FT                   /*tag=  y
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24206,g)
FT                   /*tag=  z
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24774,a)
FT                   /*tag=  aa
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (24869,c)
FT                   /*tag=  ab
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (25768,a)
FT                   /*tag=  ac
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (26697,t)
FT                   /*tag=  ad
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (26697,c)
FT                   /*tag=  ae
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (28359,c)
FT                   /*tag=  af
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (28470..28471,t)
FT                   /*tag=  ag
FT                   /note= "a single nucleotide polymorphism (SNP) can
FT                          result in a deletion at this position"
FT   variation       replace (29781,g)
FT                   /*tag=  ah
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (30182,a)
FT                   /*tag=  ai
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (31772,t)
FT                   /*tag=  aj
FT                   /note= "SNP, single nucleotide polymorphism"
FT   variation       replace (31936,g)
FT                   /*tag=  ak
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            32333..32467
FT                   /*tag=  al
FT                   /number= 7
FT   intron          32468..36825
FT                   /*tag=  am
FT                   /number= 7
FT   exon            36826..36948
FT                   /*tag=  an
FT                   /product= 8
FT   intron          36949..38129
FT                   /*tag=  ao
```

```
FT                   /number= 8
FT   exon            38130..38175
FT                   /*tag=  ap
FT                   /number= 9
FT   intron          38176..39984
FT                   /*tag=  aq
FT                   /number= 9
FT   exon            39985..40088
FT                   /*tag=  ar
FT                   /number= 10
FT   intron          40089..42455
FT                   /*tag=  as
FT                   /number= 10
FT   exon            42456..42577
FT                   /*tag=  at
FT                   /number= 11
FT   intron          42578..44422
FT                   /*tag=  au
FT                   /number= 11
FT   variation       replace (42767..42767,c)
FT                   /*tag=  av
FT                   /note= "a single nucleotide polymorphism (SNP) can
FT                          result in a deletion at this position"
FT   exon            44423..44691
FT                   /*tag=  aw
FT                   /number= 12
FT   intron          44692..47818
FT                   /*tag=  ax
FT                   /number= 12
FT   exon            47819..47897
FT                   /*tag=  ay
FT                   /number= 13
FT   intron          47898..50266
FT                   /*tag=  az
FT                   /number= 13
FT   variation       replace (48839,t)
FT                   /*tag=  ba
FT                   /note= "SNP, single nucleotide polymorphism"
FT   exon            50267..50651
FT                   /*tag=  bb
FT                   /number= 14
FT   variation       replace (52265,g)
FT                   /*tag=  bc
FT                   /note= "SNP, single nucleotide polymorphism"
XX
PN   US2002142951-A1.
XX
PD   03-OCT-2002.
XX
PF   28-MAR-2001; 2001US-0818264.
XX
PR   28-MAR-2001; 2001US-0818264.
XX
PA   (WEBS/) WEBSTER M.
PA   (BEAS/) BEASLEY E M.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
XX
PI   Webster M,  Beasley EM,  Ketchum KA,  Di Francesco V;
XX
```

```
  Query Match             4.7%;  Score 47;  DB 25;  Length 53226;
  Best Local Similarity   78.9%;  Pred. No. 0.33;
  Matches   56;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
              || ||| || ||||||| |||||| |    |||||||| |||||||||||||| || |||
Db      29825 GTTTTTAGAATTCATGCTAAATGAATAGATTTTAGCTACTCTTGCCACAAAAACAACAAA 29884

Qy        982 AAAAAAAAAAA 992
              |||| |    ||
Db      29885 AAAAGAGGTAA 29895
```

```
RESULT 25
AAX83003/c
ID   AAX83003 standard; DNA; 87350 BP.
XX
AC   AAX83003;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Human WRN genomic sequence.
XX
KW   Human; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW   recessive disorder; phenotype; ss.
XX
OS   Homo sapiens.
XX
PN   WO9724435-A1.
XX
PD   10-JUL-1997.
XX
PF   30-DEC-1996;   96WO-US20785.
XX
PR   12-APR-1996;   96US-0632175.
PR   29-DEC-1995;   95US-0009409.
PR   29-DEC-1995;   95US-0580539.
PR   30-JAN-1996;   96US-0010835.
PR   30-JAN-1996;   96US-0594242.
XX
PA   (DARW-) DARWIN MOLECULAR CORP.
PA   (OSHI/) OSHIMA J.
XX
PI   Fu Y, Mulligan J, Oshima J, Schellenberg GD, Yu C;
XX
DR   WPI; 1997-363671/33.
XX
PT   Isolated nucleic acid molecule encoding the WRN gene product -
PT   useful for detection and treatment of Werner's syndrome, and related
PT   diseases
XX
PS   Claim 1; Fig 5A-U; 153pp; English.
XX
CC   This sequence represents the genomic region containing the coding
CC   sequence for the human WRN gene which encodes a protein related to
CC   Werner's syndrome.  The products can be used for the detection and
CC   treatment of Werner's syndrome (WS), an autosomal recessive disorder
CC   with a complex phenotype, as well as related diseases.
XX
SQ   Sequence 87350 BP; 25621 A; 16221 C; 17012 G; 28450 T; 46 other;

  Query Match             4.7%;  Score 47;  DB 18;  Length 87350;
  Best Local Similarity   78.9%;  Pred. No. 0.39;
  Matches   56;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
              ||||||| ||| ||| ||||||||| ||| |||||||| |||||||||||| ||||||
Db      18894 GTATTTGGGATTCCTGCTAAATGAGTGGATTTCAGCTGCTTTTGCCACAAAAACAAAAAA 18835

Qy        982 AAAAAAAAAAA 992
              |     || ||
Db      18834 ATGGTTAACAA 18824


RESULT 26
AAS68336/c
ID   AAS68336 standard; cDNA; 636 BP.
XX
AC   AAS68336;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #4140.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG04149.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 4140; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 636 BP; 103 A; 246 C; 180 G; 107 T; 0 other;

  Query Match             4.5%;  Score 44.4;  DB 23;  Length 636;
  Best Local Similarity   88.9%;  Pred. No. 0.33;
  Matches   48;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         57 CGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGCTGCCAGCAGC 110
              |||||||||||||||||||||||||||||||||||||||||||||| ||    || |
Db        263 CGAGGCATGGGGCTCCCTGGGCTGTTCTGCTTGGCCGTGCTGGGTGAGTCCACC 210


RESULT 27
ABZ80229
ID   ABZ80229 standard; DNA; 249999 BP.
XX
AC   ABZ80229;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human tramdorin gene region genomic DNA SEQ ID NO:26.
XX
KW   Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
KW   central nervous system disorder; CNS disorder; multiple sclerosis;
KW   nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramd;
```

```
KW   tramdorin; human; chromosome 5; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003016502-A2.
XX
PD   27-FEB-2003.
XX
PF   21-AUG-2002; 2002WO-US26637.
XX
PR   21-AUG-2001; 2001US-313907P.
PR   21-AUG-2002; 2002US-0225810.
XX
PA   (MCLA-) MCLAUGHLIN RES INST.
XX
PI   Bermingham JR;
XX
DR   WPI; 2003-278567/27.
XX
PT   New nucleic acid sequence encoding tramdorins, e.g. mouse tramd 1,
PT   mouse tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd
PT   3 or rat tramd 1, useful for treating CNS, e.g. stroke, multiple
PT   sclerosis, trauma, neuropathic pain -
XX
PS   Example 6; Fig 9; 177pp; English.
XX
CC   The present invention describes an isolated nucleic acid sequence
CC   comprising a cDNA sequence encoding mouse tramdorin (tramd) 2, mouse
CC   tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or
CC   the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is
CC   located to chromosome 11, whereas human tramd 1 is located to chromosome
CC   5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic
CC   and cerebroprotective activities, and can be used in gene therapy. The
CC   nucleic acid sequences are useful for diagnosing and treating central
CC   nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
CC   neuropathic pain, stroke or trauma, and non-CNS disorders. The present
CC   sequence represents the genomic sequence of the human tramdorin gene
CC   region, which is given in the exemplification of the present invention.
XX
SQ   Sequence 249999 BP; 75050 A; 54012 C; 51931 G; 67805 T; 1201 other;

  Query Match             4.4%;  Score 44;  DB 25;  Length 249999;
  Best Local Similarity   77.9%;  Pred. No. 3.1;
  Matches   53;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        922 GTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAA 981
              ||||||  | ||| ||  ||||||||   ||||||||||||||||||||||||| |||
Db      75447 GTATTTAAAATTCCTGTGAAATGAGTAAATTTTAGCTGCTCTTGCCACAAAAAAGAAACC 75506

Qy        982 AAAAAAAA 989
              ||| ||
Db      75507 CCAAATAA 75514


RESULT 28
ABT07673
ID   ABT07673 standard; cDNA; 560 BP.
XX
AC   ABT07673;
XX
DT   14-NOV-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 152.
XX
KW   Human; breast specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200264611-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US04197.
XX
PR   13-FEB-2001; 2001US-268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Hu P,  Recipon H,  Karra K,  Cafferkey R;
PI   Sun Y,  Liu C;
XX
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer
PT   and non-cancerous disease states in breast tissue, and in gene therapy
PT   -
XX
PS   Claim 1; Page 269; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins. These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a coding sequence of the invention.
XX
SQ   Sequence 560 BP; 173 A; 109 C; 139 G; 139 T; 0 other;

  Query Match             4.4%;  Score 43.4;  DB 24;  Length 560;
  Best Local Similarity   76.8%;  Pred. No. 0.56;
  Matches   53;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy        924 ATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAA 983
              || |||  ||||| | ||||||||   |||||||| ||||||||||||| ||| |||||||
Db        408 ATATGGGATTCATACTAAATGAGTAGATTTTAGCTTCTCTTGCCACAATAAACAAAAAAA 467

Qy        984 AAAAAAAAA 992
              |    | ||
Db        468 ATGCCACAA 476


RESULT 29
ABT07672
ID   ABT07672 standard; cDNA; 509 BP.
XX
AC   ABT07672;
XX
DT   14-NOV-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 151.
XX
KW   Human; breast specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200264611-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US04197.
XX
PR   13-FEB-2001; 2001US-268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S,  Macina RA,  Hu P,  Recipon H,  Karra K,  Cafferkey R;
PI   Sun Y,  Liu C;
XX
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer
PT   and non-cancerous disease states in breast tissue, and in gene therapy
PT   -
XX
```

```
PS   Claim 1; Page 268-269; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins. These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a coding sequence of the invention.
XX
SQ   Sequence 509 BP; 151 A; 109 C; 123 G; 126 T; 0 other;

  Query Match             4.3%;  Score 43;  DB 24;  Length 509;
  Best Local Similarity   77.6%;  Pred. No. 0.68;
  Matches   52;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy        926 TTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAA 985
              ||||  |||||  | ||||||||   |||||||| |||||||||||| ||| ||||||||
Db        359 TTGGGATTCATACTAAATGAGTAGATTTTAGCTTCTCTTGCCACAATAAACAAAAAAAAT 418

Qy        986 AAAAAAA 992
                | ||
Db        419 GCCACAA 425


RESULT 30
AAL38336/c
ID   AAL38336 standard; DNA; 143899 BP.
XX
AC   AAL38336;
XX
DT   15-AUG-2002  (first entry)
XX
DE   Genomic sequence encoding a human NgR2 protein.
XX
KW   Cerebroprotective; neuroprotective; cytostatic; Nogo receptor homologue;
KW   NgR2; NgR3; axonal growth; central nervous system; CNS; cerebral injury;
KW   spinal cord injury; stroke; demyelinating disease; multiple sclerosis;
KW   monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;
KW   multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;
KW   Alexander's disease; Canavan's disease; metachromatic leukodystrophy;
KW   Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;
KW   transgenic animal; unregulated cellular growth; cancer; tumour; human;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200229059-A2.
XX
PD   11-APR-2002.
XX
PF   06-OCT-2001; 2001WO-US31488.
XX
PR   06-OCT-2000; 2000US-238361P.
XX
PA   (UYYA ) UNIV YALE.
PA   (BIOJ ) BIOGEN INC.
XX
PI   Strittmatter SM,  Cate RL,  Sah DWY;
XX
DR   WPI; 2002-416677/44.
XX
PT   Novel Nogo receptor homolog polypeptide, NgR2 or NgR3, useful for
PT   treating central nervous system disorder, cerebral injury, spinal cord
PT   injury, stroke, and demyelinating diseases  -
XX
PS   Example 2; Page 176-214; 277pp; English.
XX
CC   The invention relates to a Nogo receptor homologue polypeptide, NgR2 or
CC   NgR3, comprising a 50 amino acid LRRCT sequence, a 284 amino acid NTLRRCT
CC   sequence, or a 420, 461 or 392 amino acid sequence, all given in the
CC   specification. The NgR3 protein or its binding antibody is useful for
CC   decreasing inhibition of axonal growth of a central nervous system (CNS)
CC   neuron, by contacting the neuron NgR3 or its antibody, and for treating
CC   CNS disease, disorder or injury. NgR3 or a vector comprising NgR3 is
CC   useful for treating cerebral injury, spinal cord injury, stroke,
CC   demyelinating diseases, e.g. multiple sclerosis, monophasic
CC   demyelination, encephalomyelitis, multifocal leukoencephalopathy,
CC   panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,
CC   Alexander's disease, Canavan's disease, metachromatic leukodystrophy and
CC   Krabbe's disease. NgR3 is useful for inducing an immune response in a
CC   mammal against NgR3, as a bait protein in a two-hybrid or three-hybrid
CC   assay, and as a research tool for identification, characterisation and
CC   purification of interacting, regulatory proteins. The nucleotide
CC   sequences of the invention are useful for screening for RFLP associated
CC   with certain disorders, for genetic mapping, and for gene therapy. The
CC   vector containing NgR3 is useful for producing non-human transgenic
CC   animals. The NgR3 binding antibody is useful for isolating and purifying
CC   NgR3, for localisation and/or quantitation of NgR3, and for diagnostic
CC   and therapeutic purposes. The sequences of the invention, vectors and
CC   antibodies are useful for treating or preventing unregulated cellular
CC   growth such as cancer and tumour growth. This polynucleotide sequence
CC   represents the genomic sequence encoding a human NgR2 protein of the
CC   invention.
XX
SQ   Sequence 143899 BP; 36346 A; 35277 C; 35318 G; 35657 T; 1301 other;

  Query Match             4.3%;  Score 42.8;  DB 24;  Length 143899;
  Best Local Similarity   56.1%;  Pred. No. 5;
  Matches  101;  Conservative    0;  Mismatches   77;  Indels    2;  Gaps    1;

Qy        804 GGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAAT 863
              ||| ||   |||| | || |   | |||||  |||||  |||| |||||    | || ||
Db      92710 GGGTGGGAGAGGAGATGGGAGATGCAGGTCGAAGGACAAAAAGGAGCAGATGAGCAGGAT 92651

Qy        864 GAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGT 923
              ||||  | ||        | |   |     | ||   | |    |||  |   |    ||
Db      92650 GAACAGGCCTAGAGATCTAATGCACAGCACGAAGACTGTAGGTCATCAAACTGCATTGTG 92591

Qy        924 ATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAA 983
              || ||| ||  | | |||  |||   ||||| ||||||||| |||||||  |   | |||
Db      92590 TTTGGGACTT--TCCTAAACAAGTAGATTTTAACTGCTCTTGTCACAAAACAGGTAGAAA 92533


RESULT 31
AAF21712/c
ID   AAF21712 standard; DNA; 752 BP.
XX
AC   AAF21712;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human breast and ovarian cancer associated antigen gene SEQ ID 99.
XX
KW   Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW   nootropic; neurpprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
KW   Addison's disease; allergy; autoimmune haemolytic anaemia;
KW   autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW   multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW   cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055173-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05881.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-611515/58.
DR   P-PSDB; AAB58809.
```

XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 1; Page 555-556; 1299pp; English.
XX
CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 752 BP; 149 A; 275 C; 219 G; 104 T; 5 other;

Query Match 4.2%; Score 41.8; DB 21; Length 752;
Best Local Similarity 51.3%; Pred. No. 1.5;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 410 GATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCA 469
Db 260 GCTGGGCTGGCAGATGCAGTCCACGGCGCTGTCGTAGCGCAGGAGCACGTGGCCAGGACC 201

Qy 470 GGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACC 529
Db 200 CCGCAGAGGGGGCGCCCGCAGACCGCGGCGGCCCTGCCCGTAGTCGCCAGTCAGCACCCG 141

Qy 530 TATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCA 589
Db 140 TGCCACGAACACCGCCTTATGGCCATCGGCGTTGGGGGGCGAGTAGCGGTCCTGCACCGA 81

Qy 590 CAGGCAGCC 598
Db 80 CAGGGAGGC 72

RESULT 32
AAC77293/c
ID AAC77293 standard; cDNA; 1417 BP.
XX
AC AAC77293;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2848 polynucleotide sequence SEQ ID NO:5695.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB43084.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 4858-4859; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1417 BP; 211 A; 499 C; 473 G; 234 T; 0 other;

Query Match 4.2%; Score 41.8; DB 21; Length 1417;
Best Local Similarity 51.3%; Pred. No. 1.9;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 410 GATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCA 469
Db 987 GCTGGGCTGGCAGATGCAGTCCACGGCGCTGTCGTAGCGCAGGAGCACGTGGCCAGGACC 928

Qy 470 GGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACC 529
Db 927 CCGCAGAGGGGGCGCCCGCAGACCGCGGCGGCCCTGCCCGTAGTCGCCAGTCAGCACCCG 868

Qy 530 TATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCA 589
Db 867 TGCCACGAACACCGCCTTATGGCCATCGGCGTTGGGGGGCGAGTAGCGGTCCTGCACCGA 808

Qy 590 CAGGCAGCC 598
Db 807 CAGGGAGGC 799

```
RESULT 33
AAC77637
ID   AAC77637 standard; cDNA; 2059 BP.
XX
AC   AAC77637;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:31.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; AAB43428.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 633-634; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 2059 BP; 505 A; 548 C; 552 G; 451 T; 3 other;

  Query Match             4.2%;  Score 41.8;  DB 21;  Length 2059;
  Best Local Similarity   61.5%;  Pred. No. 2.2;
  Matches   67;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy        884 CACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAAT 943
              || | |||||  ||||  | | || || ||  ||| || | |        | |  | | |
Db       1945 CAGGCCAGAGTGCTGCCAGTCCTCTGCATGTTCTGCTCTTGTGCCCTTCTGAGCCCACAA 2004

Qy        944 GAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
              |  |  |    ||  |||||| | ||||||||||||||||||||||||
Db       2005 TAAAGGCTGAGCTCTTATCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAA 2053


RESULT 34
ABX63699/c
ID   ABX63699 standard; cDNA; 2672 BP.
XX
AC   ABX63699;
XX
DT   26-FEB-2003  (first entry)
XX
DE   Human cDNA #699 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis;
KW   cardiant; hypotensive; antidiabetic; gynaecological; vasotropic;
KW   cerebroprotective; gene therapy; vascular disease; cancer; coronary;
KW   artery disease; hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke;
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-0044090.
XX
PR   28-JUL-2000; 2000US-222469P.
PR   08-JAN-2001; 2001US-260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue -
XX
PS   Claim 1; Page -; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that
CC   are differentially expressed in activated vascular tissue. The invention
CC   also discloses a high throughput method for detecting differentially
CC   expressed cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a
CC   high-throughput methods for detecting differential expression of one or
CC   more cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of
CC   treatment of a vascular disease, e.g. atherosclerosis, cancer, coronary
CC   artery disease, hypertension, diabetes, pre-eclampsia, ischaemia-
CC   reperfusion injury, restenosis, or stroke. The cDNAs can also be used
CC   for large-scale genetic or gene expression analysis of several new
CC   nucleic acid molecules. Antibodies to the proteins encoded by the
CC   cDNAs are useful for diagnosing pre-pathologic disorders, and chronic
CC   or acute diseases associated with abnormalities in the expression,
CC   amount or distribution of the protein. The present sequence
CC   represents a cDNA of the invention that is differentially expressed in
```

CC activated vascular tissue.
CC Note: The sequence data for this patent did not form part of the
CC specification, but was obtained in electronic format directly from USPTO
CC at http.seqdata.uspto.gov/sequence.html?DocID=20020137081.
XX
SQ Sequence 2672 BP; 432 A; 880 C; 907 G; 453 T; 0 other;

Query Match 4.2%; Score 41.8; DB 25; Length 2672;
Best Local Similarity 51.3%; Pred. No. 2.4;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 410 GATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCA 469
Db 2220 GCTGGGCTGGCAGATGCAGTCCACGGCGCTGTCGTAGCGCAGGAGCACGTGGCCAGGACC 2161

Qy 470 GGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACC 529
Db 2160 CCGCAGAGGGGGCGCCCGCAGACCGCGGCGGCCCTGCCCGTAGTCGCCAGTCAGCACCCG 2101

Qy 530 TATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCA 589
Db 2100 TGCCACGAACACCGCCTTATGGCCATCGGCGTTGGGGGGCGAGTAGCGGTCCTGCACCGA 2041

Qy 590 CAGGCAGCC 598
Db 2040 CAGGGAGGC 2032

RESULT 35
AAH15559/c
ID AAH15559 standard; cDNA; 3471 BP.
XX
AC AAH15559;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:13847.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13847; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 3471 BP; 568 A; 1088 C; 1233 G; 582 T; 0 other;

Query Match 4.2%; Score 41.8; DB 22; Length 3471;
Best Local Similarity 51.3%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 410 GATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCA 469
Db 3027 GCTGGGCTGGCAGATGCAGTCCACGGCGCTGTCGTAGCGCAGGAGCACGTGGCCAGGACC 2968

Qy 470 GGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACC 529
Db 2967 CCGCAGAGGGGGCGCCCGCAGACCGCGGCGGCCCTGCCCGTAGTCGCCAGTCAGCACCCG 2908

Qy 530 TATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCA 589
Db 2907 TGCCACGAACACCGCCTTATGGCCATCGGCGTTGGGGGGCGAGTAGCGGTCCTGCACCGA 2848

Qy 590 CAGGCAGCC 598
Db 2847 CAGGGAGGC 2839

RESULT 36
ABL53715/c
ID ABL53715 standard; cDNA; 3541 BP.
XX
AC ABL53715;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human poly(ADP-ribose) polymerase p150 cDNA.
XX
KW Poly(ADP-ribose) polymerase; PARP; p150; human; enzyme; diabetes;
KW cerebral ischaemia; Alzheimer's disease; Parkinson's disease;
KW systemic lupus erythematosus; arthritis; tumour; therapy;
KW diagnosis; cytostatic; vasotropic; antidiabetic; nootropic;
KW neuroprotective; antiparkinsonian; antiarthritic; dermatological;
KW immunosuppressive; antiinflammatory; expressed sequence tag; EST;
KW gene; chromosome 8; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 122..3199
FT /*tag= a
FT /product= "Poly(ADP-ribose) polymerase"
XX
PN WO200222792-A2.
XX
PD 21-MAR-2002.
XX
PF 11-SEP-2001; 2001WO-EP10494.
XX
PR 12-SEP-2000; 2000EP-0119849.

PR 13-JUN-2001; 2001EP-0113577.
XX
PA (LUES/) LUESCHER B.
PA (AUST/) AUSTEN M.
XX
PI Luescher B, Austen M;
XX
DR WPI; 2002-339867/37.
DR P-PSDB; ABB75711.
XX
PT New polynucleotide encoding poly(ADP-ribose) polymerase useful as a
PT hybridization probe, and for preparing a pharmaceutical composition for
PT preventing and/or treating tumors -
XX
PS Claim 1; Fig 3; 76pp; English.
XX
CC The present sequence is that of cDNA encoding a human protein,
CC termed p150, having poly(ADP-ribose) polymerase (PARP) activity.
CC p150 protein was isolated from a native c-Myc complex. Predicted
CC DNA sequences from p150 tryptic peptides were used to screen
CC expressed sequence tag (EST) databases. The present, full-length
CC cDNA was assembled from 2 EST clones. The p150 gene was assigned
CC to chromosome 8 band q24 in close proximity to the c-Myc gene,
CC suggesting a role for the gene in Burkitt's lymphoma, and breast
CC and prostate carcinomas. The p150 protein may participate in
CC activating Myc/Max-driven transcription. It has the capacity to
CC stimulate Myc-mediated transcriptional activation, and the
CC capacity to inhibit cell growth, especially cell proliferation,
CC and/or to inhibit Myc/Ras-mediated transformation of cells. p150
CC shows ubiquitous expression, and p150 transcription is elevated in
CC the thymus. The invention provides p150 nucleic acids and
CC polypeptides, expression vectors, host cells, antibodies,
CC pharmaceutical compositions and methods for treating disorders
CC associated with aberrant regulation of cellular behaviour. Use of
CC a p150 antagonist for the treatment of cerebral ischaemia, diabetes,
CC Alzheimer's disease, Parkinson's disease, systemic lupus erythematosus
CC or arthritis, or for the prevention and/or treatment of tumours, is
CC claimed. Also claimed is a diagnostic composition comprising a p150
CC nucleic acid, protein or antibody.
XX
SQ Sequence 3541 BP; 592 A; 1108 C; 1251 G; 590 T; 0 other;

Query Match 4.2%; Score 41.8; DB 24; Length 3541;
Best Local Similarity 51.3%; Pred. No. 2.6;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```
Qy        410 GATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAACTTCACTCTGCA 469
              | ||  |||| || ||  ||||| | |  ||||  |||   ||  ||  |   |     |
Db       3079 GCTGGGCTGGCAGATGCAGTCCACGGCGCTGTCGTAGCGCAGGAGCACGTGGCCAGGACC 3020

Qy        470 GGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTCGGGCAGCCCACC 529
              ||||||||  | || |||   |  |  |  |||||  |  |||  | | |||| | |
Db       3019 CCGCAGAGGGGGCGCCCGCAGACCGCGGCGGCCCTGCCCGTAGTCGCCAGTCAGCACCCG 2960

Qy        530 TATCACCAACAGCCTGATCGGGAAGGATGGGCAGGTCCACCTGCAGCAGAGACCATGCCA 589
              |  ||| |||| |     |  ||       | |  ||     |  | |||  |        | |
Db       2959 TGCCACGAACACCGCCTTATGGCCATCGGCGTTGGGGGGCGAGTAGCGGTCCTGCACCGA 2900

Qy        590 CAGGCAGCC 598
              |||| || |
Db       2899 CAGGGAGGC 2891
```

RESULT 37
ABQ99652
ID ABQ99652 standard; DNA; 38736 BP.
XX
AC ABQ99652;
XX
DT 12-NOV-2002 (first entry)
XX
DE Human membrane spanning 4-domain family, subfamily A genomic sequence.
XX
KW Human; membrane spanning 4-domain A; cytostatic; antiallergic; MS4A;
KW gene therapy; atopic disorder; non-Hodgkin's lymphoma;
KW Hodgkin's lymphoma; allergenic disease; gene; ss; chromosome 11q12-13.1.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	7631..12690
FT		/*tag= a
FT		/product= "MS4A"
FT		/note= "Contains 2 introns"
FT	exon	7631..7729
FT		/*tag= b
FT		/number= 1
FT	intron	7730..10475
FT		/*tag= c
FT		/number= 1
FT		/cons_splice= (5'site:NO,3'site:YES)
FT	exon	10476..10682
FT		/*tag= d
FT		/number= 2
FT	intron	10683..12600
FT		/*tag= e
FT		/number= 2
FT	exon	12601..12687
FT		/*tag= f
FT		/number= 3

XX
PN WO200262946-A2.
XX
PD 15-AUG-2002.
XX
PF 10-DEC-2001; 2001WO-US48437.
XX
PR 08-DEC-2000; 2000US-254362P.
PR 20-FEB-2001; 2001US-270057P.
XX
PA (UYDU-) UNIV DUKE.
XX
PI Tedder TF,. Liang YH;
XX
DR WPI; 2002-657530/70.
XX
PT New membrane spanning 4-domain A (MS4A) genes and polypeptides, useful
PT for generating animal models of atopic disorders, for drug screening,
PT or for treating (non-) Hodgkin's lymphoma, or allergenic or atopic
PT disorders in e.g. humans -
XX
PS Claim 18; Page 304-325; 450pp; English.
XX
CC The invention relates to novel membrane spanning 4-domain A (MS4A)
CC nucleic acid and polypeptide molecules, comprising human and mouse MS4A.
CC The polypeptides of the invention have cytostatic and antiallergic
CC activity. The polynucleotides may have a use in gene therapy. The MS4A
CC nucleic acids and polypeptides are useful for generating animal (e.g.
CC mouse) models of atopic disorders, or for drug discovery screens. These
CC are also useful for treating (non-)Hodgkin's lymphoma, allergenic
CC diseases, atopic disorders or other MS4A-related conditions. The present
CC sequence represents a human MS4A genomic region.
XX
SQ Sequence 38736 BP; 11258 A; 7734 C; 8078 G; 11665 T; 1 other;

Query Match 4.1%; Score 41; DB 24; Length 38736;
Best Local Similarity 76.9%; Pred. No. 9.1;
Matches 50; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy        920 TCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
              | ||||||||  ||  ||| ||||||  |   |||||||||||||||||||| | | |||
Db       1630 TTGTATTTGGCATTTTTGCTAAATGACTAGATTTTAGCTGCTCTTGCCACACATAGAAAT 1689

Qy        980 AAAAA 984
              | |||
Db       1690 ACAAA 1694
```

```
RESULT 38
ABN97975
ID   ABN97975 standard; DNA; 44100 BP.
XX
AC   ABN97975;
XX
DT   01-AUG-2002  (first entry)
XX
DE   Human retroviral sequence RAM75.
XX
KW   Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW   multiple sclerosis; ds.
XX
OS   Human retrovirus.
XX
PN   WO9967395-A1.
XX
PD   29-DEC-1999.
XX
PF   23-JUN-1999;   99WO-FR01513.
XX
PR   23-JUN-1998;   98FR-0007920.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Alliel PM,  Perin J,  Rieger F;
XX
DR   WPI; 2000-160587/14.
XX
PT   New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT   used for diagnosis, treatment and prevention of autoimmune and
PT   neurological diseases  -
XX
PS   Claim 15; Fig 14; 225pp; French.
XX
CC   The present invention relates to new nucleic acid sequences of human
CC   endogenouse retrovirus, HERV-7q, which is located on chromosome 7q.
CC   Regulatory elements associated with HERV-7q may alter expression of other
CC   genes (even remote genes) on the same chromosome, inducing immunological
CC   and/or neurological changes (which may be pathological or protective/
CC   curative). HERV-7q peptides can be used to improve efficiency of the
CC   immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC   sequences can be used in immunogenic or vaccinating compositions, for
CC   protection against autoimmune diseases, particularly multiple sclerosis.
CC   The peptides may also be used (by sequence comparison) to detect/identify
CC   endogenous retroviruses that are abnormally expressed in cancer,
CC   neuropathologies or other autoimmune diseases. The present sequence was
CC   used to illustrate the invention.
XX
SQ   Sequence 44100 BP; 13051 A; 8074 C; 8534 G; 14441 T; 0 other;

  Query Match             4.0%;  Score 39.6;  DB 21;  Length 44100;
  Best Local Similarity   77.4%;  Pred. No. 21;
  Matches   48;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy        920 TCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAA 979
              | ||||| ||  ||||||  ||||||||   ||||| || ||||| |||||||||| | ||
Db      10427 TTGTATTGGGGATTCATGTTAAATGAGTAGATTTTAACTACTCTTACCACAAAAACACAA 10486

Qy        980 AA 981
              ||
Db      10487 AA 10488


RESULT 39
ABN22420
ID   ABN22420 standard; cDNA; 276 BP.
XX
AC   ABN22420;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:13317.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP06668.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 13317; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 276 BP; 43 A; 105 C; 86 G; 40 T; 2 other;

  Query Match             4.0%;  Score 39.4;  DB 24;  Length 276;
  Best Local Similarity   50.8%;  Pred. No. 4.4;
  Matches   94;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        338 TCCAGACCTGCTCACCTACTTCTGCCGGGCGTCCTCCACCTCAGGTGCCCATGTGGACAG 397
              |||  |||||||        ||  ||  |    | |||||    |   ||  |||  ||
Db          9 TCCCCACCTGCTCGAGCCGCTCGACCTCGGCCACACCACCCTGCGCAACCGCGTGATCAT 68

Qy        398 TGCCAGGCTACAGATGCACTGGGAGCTGTGGTCCAAGCCAGTGTCTGAGCTGCGGGCCAA 457
```

```
Db         69 GGGCTCGATGCACACCGGCATGGAGGATCGTGCCAAGCACCTGCCCGAGCTGGCGGCCTA 128

Qy        458 CTTCACTCTGCAGGACAGAGGGGCAGGCCCCAGGGTGGAGATGATCTGCCAGGCGTCCTC 517

Db        129 CTTCGCCGAGCGGGCCAAGGGTGGCGTGGCCCTGTCGGTCACCGGCGGCTACCGCCCAAC 188

Qy        518 GGGCA 522

Db        189 TGGCA 193
```

RESULT 40
AAX80672
ID AAX80672 standard; cDNA; 3285 BP.
XX
AC AAX80672;
XX
DT 22-OCT-1999 (first entry)
XX
DE Clone nn103_4 encoding secreted protein-nn103_4.
XX
KW Secreted protein; cytokine; cell proliferation; immune stimulation;
KW vaccine; immune suppression; haematopoiesis; tissue growth; activin;
KW inhibin; chemotaxis; chemokinesis; haemostasis; thrombolytic; ss;
KW receptor; ligand; anti-inflammatory; cadherin; tumour; gene therapy.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   2294..2848
FT   /*tag= a
FT   /product= "Secreted protein nn103_4"
XX
PN WO9928335-A1.
XX
PD 10-JUN-1999.
XX
PF 02-DEC-1998; 98WO-US25512.
XX
PR 30-NOV-1998; 98US-0203106.
PR 04-DEC-1997; 97US-0067454.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Clark HF, Collins-Racie LA, Evans C;
PI Fechtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Steininger RJ, Treacy M, Wong GG;
XX
DR WPI; 1999-385352/32.
DR P-PSDB; AAY26037.
XX
PT New polynucleotides encoding secreted human proteins
XX
PS Claim 28; Pages 111-112; 124pp; English.
XX
CC The present sequence is a known clone nn103_4 (deposited as ATCC
CC 98600) isolated from human foetal kidney cDNA library. It encodes
CC secreted protein nn103_4. Recombinant secreted proteins can be produced
CC by transforming host cells and culturing them under suitable conditions.
CC The polynucleotide and protein are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Some predicted
CC biological activities include cytokine and cell proliferation/
CC differentiation activity, immune stimulating (e.g. as vaccines) or
CC suppressing activity, haematopoiesis regulating activity, tissue growth
CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
CC hemostatic and thrombolytic activity, receptor/ligand activity, anti-
CC inflammatory activity, cadherin/tumour invasion suppressor activity, and
CC tumour inhibition activity. The polynucleotide can be used for gene
CC therapy.
XX
SQ Sequence 3285 BP; 756 A; 813 C; 705 G; 1011 T; 0 other;

```
  Query Match             4.0%;  Score 39.4;  DB 20;  Length 3285;
  Best Local Similarity   72.5%;  Pred. No. 10;
  Matches   66;  Conservative    0;  Mismatches   21;  Indels    4;  Gaps    1;

Qy        891 GAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTG 950

Db       3049 GAGGAATGTAGGTAATAAAATTGCACTGTT----TTGGGAGTTCCTGCTAAATGACTAGA 3104

Qy        951 TTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981

Db       3105 CTTCAGCTGCTCTTGCCACAAAATCCTAAAA 3135
```

RESULT 41
AAS59222
ID AAS59222 standard; cDNA; 3285 BP.
XX
AC AAS59222;
XX
DT 16-JAN-2002 (first entry)
XX
DE Human cDNA encoding a secreted protein nn103_4.
XX
KW Human; secreted protein; ss; antiinflammatory; immunosuppressive;
KW nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;
KW cytostatic; antidiabetic; virucide; antiinfertility; anticonvulsant;
KW vasotropic; antiparkinsonian; immunostimulant; dermatological;
KW antirheumatic; antitumor; antiulcer; osteopathic; tranquiliser;
KW cerebroprotective; cytokine; cell proliferation; cell differentiation;
KW immune deficiency; severe combined immunodeficiency; SCID; tumour;
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;
KW food supplement.
XX
OS Homo sapiens.
XX
PN WO200175068-A2.
XX
PD 11-OCT-2001.
XX
PF 22-MAR-2001; 2001WO-US09369.
XX
PR 30-MAR-2000; 2000US-0539330.
PR 04-DEC-2000; 2000US-0729674.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;
PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;
PI Clark H, Fechtel K, Merberg D;
XX
DR WPI; 2001-639363/73.
DR P-PSDB; AAU39004.
XX
PT Secreted human proteins, useful as vaccine for treating various
PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and
PT nervous system disorders (e.g. stroke) -
XX
PS Disclosure; Page 476-477; 619pp; English.
XX
CC The invention relates to novel human secreted proteins, the nucleic
CC acids encoding them. The protein may exhibit cytokine, cell proliferation
CC or cell differentiation activity or may induce production of other
CC cytokines in certain cell populations and may exhibit immune stimulating
CC or immune suppressing activity, which is useful for the treatment of
CC various immune deficiencies and disorders e.g. severe combined
CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,
CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation. The proteins are also useful in the treatment of diseases

CC and disorders including tissue, skin and organ transplantation and in
CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,
CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,
CC in the treatment of burns, incisions and ulcers; as well as in treatment
CC of periodontal disease, osteoporosis or osteoarthritis, mediated by
CC inflammatory processes, diseases of the peripheral nervous system,
CC Alzheimer's, Parkinson's disease, Huntington's disease,
CC amylotrophic lateral sclerosis, and Shy-Drager syndrome, infections,
CC infarction of cardiac and central nervous system vessel e.g. stroke,
CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The
CC protein, having activin- or inhibin-related activities is useful as a
CC contraceptive based on the ability of inhibins to decrease fertility in
CC female mammals and decrease spermatogenesis in male mammals. The
CC proteins and nucleic acids are also useful as food supplements. The
CC present sequence encodes a secreted protein of the invention.
XX
SQ Sequence 3285 BP; 756 A; 813 C; 705 G; 1011 T; 0 other;

Query Match 4.0%; Score 39.4; DB 22; Length 3285;
Best Local Similarity 72.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 21; Indels 4; Gaps 1;

Qy 891 GAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTG 950
Db 3049 GAGGAATGTAGGTAATAAAATTGCACTGTT----TTGGGAGTTCCTGCTAAATGACTAGA 3104

Qy 951 TTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
Db 3105 CTTCAGCTGCTCTTGCCACAAAATCCTAAAA 3135

RESULT 42
ABA90891
ID ABA90891 standard; cDNA; 3285 BP.
XX
AC ABA90891;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 31.
XX
KW Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerary;
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.
XX
OS Homo sapiens.
XX
PN US2001039335-A1.
XX
PD 08-NOV-2001.
XX
PF 04-DEC-2000; 2000US-0729674.
XX
PR 10-APR-1997; 97US-126425P.
PR 04-DEC-1997; 97US-067454P.
PR 20-DEC-1997; 97US-068379P.
PR 02-JAN-1998; 98US-070346P.
PR 07-JAN-1998; 98US-070643P.
PR 08-JAN-1998; 98US-070755P.
PR 13-JAN-1998; 98US-071304P.
PR 22-JAN-1998; 98US-072134P.
PR 30-JAN-1998; 98US-073095P.
PR 18-FEB-1998; 98US-075038P.
PR 30-MAR-2000; 2000US-0539330.
PR 23-NOV-1998; 98US-0197886.
XX
PA (JACO/) JACOBS K.
PA (MCCO/) MCCOY J M.
PA (LAVA/) LAVALLIE E R.
PA (COLL/) COLLINS-RACIE L A.
PA (EVAN/) EVANS C.
PA (MERB/) MERBERG D.
PA (TREA/) TREACY M.
PA (AGOS/) AGOSTINO M J.
PA (STEI/) STEININGER R J.
PA (SPAU/) SPAULDING V.
PA (WONG/) WONG G G.
PA (CLAR/) CLARK H.
PA (FECH/) FECHTEL K.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI Wong GG, Clark H, Fechtel K;
XX
DR WPI; 2002-040725/05.
DR P-PSDB; ABB55713.
XX
PT New secreted proteins and encoding polynucleotides, useful in gene
PT therapies, particularly for preventing or treating autoimmune
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,
PT stroke or inflammations -
XX
PS Disclosure; Page 198-199; 349pp; English.
XX
CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and
CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and
CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in
CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1
CC are deposited with the American Type Culture Collection (ATCC) with
CC accession number 98599. The polynucleotides and encoded polypeptides have
CC cytostatic, anti-inflammatory, immunomodulator, vulnerary,
CC neuroprotective, activin, inhibin, chemotactic, hAemostatic, thrombolytic
CC and anti-inflammatory activity and acting as cytokine modulators,
CC haematopoiesis regulators, tissue growth modulators and/or cadherin
CC suppressors. The polypeptides and polynucleotides are useful in gene
CC therapies, particularly for preventing, treating or ameliorating any of
CC the following diseases: immune deficiency and disorders; e.g. bacterial
CC or fungal infections, autoimmune disorders, cancer, systemic lupus
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell
CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or
CC osteoarthritis; central and peripheral nervous system diseases and
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis
CC or systemic inflammatory response syndrome, ischaemia-reperfusion
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus
CC foliaceus.
XX
SQ Sequence 3285 BP; 756 A; 813 C; 705 G; 1011 T; 0 other;

Query Match 4.0%; Score 39.4; DB 24; Length 3285;
Best Local Similarity 72.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 21; Indels 4; Gaps 1;

Qy 891 GAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTG 950
Db 3049 GAGGAATGTAGGTAATAAAATTGCACTGTT----TTGGGAGTTCCTGCTAAATGACTAGA 3104

Qy 951 TTTTAGCTGCTCTTGCCACAAAAAAAAAAAA 981
Db 3105 CTTCAGCTGCTCTTGCCACAAAATCCTAAAA 3135

RESULT 43
AAF65032/c
ID AAF65032 standard; cDNA; 394 BP.
XX
AC AAF65032;

XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 788.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 657-658; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 394 BP; 103 A; 90 C; 115 G; 86 T; 0 other;

Query Match 4.0%; Score 39.2; DB 22; Length 394;
Best Local Similarity 66.7%; Pred. No. 5.5;
Matches 56; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 909 GCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCA 968
Db 393 GCGCGCGCCGCTCTTGGGTGGAGAACATACAACAACACAGAGTGTGAGTGTGTTTTCCCC 334

Qy 969 CAAAAAAAAAAAAAAAAAAAAAAA 992
Db 333 CACAAAAAAAAAAAAAAAAAAGAA 310

RESULT 44
AAC98202
ID AAC98202 standard; cDNA; 1639 BP.
XX
AC AAC98202;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:212.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR P-PSDB; AAB53445.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 1; Page 636; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 1639 BP; 410 A; 398 C; 357 G; 469 T; 5 other;

Query Match 3.9%; Score 38.8; DB 21; Length 1639;
Best Local Similarity 64.4%; Pred. No. 11;
Matches 58; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 903 CCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTC 962
Db 1501 CCAGCACCATTTATGAGTCTCAAGTTTTATTATTGCAATAAAAGTGCTTTATGCCGGCTT 1560

Qy 963 TTGCCACAAAAAAAAAAAAAAAAAAAAAA 992
Db 1561 TTCTCAAAAAAAAAAAAAAAAAAAAAAAA 1590

RESULT 45
ABK33656
ID ABK33656 standard; cDNA; 2025 BP.
XX
AC ABK33656;
XX
DT 08-MAY-2002 (first entry)

XX
DE cDNA encoding human PRO protein, Seq ID No 241.
XX
KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW breast cancer; prostate tumour; rectal tumour; liver tumour;
KW pericyte cell proliferation; chondrocyte cell proliferation;
KW tumour necrosis factor-alpha; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200208288-A2.
XX
PD 31-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US21066.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220585P.
PR 25-JUL-2000; 2000US-220605P.
PR 25-JUL-2000; 2000US-220607P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220638P.
PR 25-JUL-2000; 2000US-220664P.
PR 25-JUL-2000; 2000US-220666P.
PR 26-JUL-2000; 2000US-220893P.
PR 28-JUL-2000; 2000WO-US20710.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 15-SEP-2000; 2000US-000000P.
PR 10-NOV-2000; 2000WO-US30873.
PR 28-NOV-2000; 2000US-253646P.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001WO-US17092.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX
DR WPI; 2002-172001/22.
DR P-PSDB; AAU83712.
XX
PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT useful for treating a PRO related disorder and for diagnosing tumours
PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT tumour or liver tumour -
XX
PS Claim 2; Figure 241; 359pp; English.
XX
CC The invention relates to one hundred and twenty two nucleic acids
CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
CC encode human secreted proteins. The PRO nucleic acids, polypeptides,
CC agonists and antagonists are useful for treating a PRO related disorder.
CC The PRO polypeptides are useful for diagnosing tumours, especially lung
CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
CC liver tumour. The PRO polypeptides are useful for stimulating the
CC proliferation of, or gene expression, in pericyte cells, for stimulating
CC the proliferation or differentiation of chondrocyte cells, for
CC stimulating the release of tumour necrosis factor-alpha from human blood,
CC for stimulating or inhibiting the proliferation of normal human dermal
CC fibroblast cells. The PRO polypeptide may also be used as molecular
CC weight markers and for tissue typing. The PRO nucleic acids have
CC applications in molecular biology, including use as hybridisation probes,
CC and in chromosome and gene mapping. ABK33536-ABK33657 represent human
CC PRO protein coding sequences of the invention.
XX
SQ Sequence 2025 BP; 471 A; 592 C; 584 G; 378 T; 0 other;

Query Match 3.9%; Score 38.8; DB 24; Length 2025;
Best Local Similarity 60.4%; Pred. No. 12;
Matches 64; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 887 GGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAG 946
Db 1886 GGTGGTGGGCCAGCGGACCTCAGGGCGGAGGCTTCCCACGGGGAGGCAGGAAGAAATAAA 1945

Qy 947 TGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAA 992
Db 1946 GGTCTTTGGCTTTCTCCAGGCAAAAAAAAAAAAAAAAAAAAAAAA 1991

RESULT 46
ABS69895/c
ID ABS69895 standard; DNA; 56737 BP.
XX
AC ABS69895;
XX
DT 21-NOV-2002 (first entry)
XX
DE Human hypoxanthine-guanine phosphoribosyltransferase (HPRT)gene.
XX
KW Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;
KW adenosine deaminase deficiency; severe combined immune deficiency; PAH;
KW beta-chain; haemoglobin gene; beta-thalassaemia; sickle cell disease;
KW low density lipoprotein gene; familial hypercholesterolaemia;
KW hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;
KW phenylalanine hydroxylase gene; gene therapy; phenylketonuria;
KW dystrophin gene; muscular dystrophy; cystic fibrosis; immunostimulant;
KW human cystic fibrosis transmembrane conductance regulator gene;
KW antianaemic; antilipaemic; nootropic; cytostatic; dermatological;
KW human; alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;
KW lysosomal arylsulfatase A; omithine transcarbamylase; ARSA; OTC; NP;
KW purin nucleoside phosphorylase; gene; ds.
XX
OS Homo sapiens.
XX
PN US2002102731-A1.
XX
PD 01-AUG-2002.
XX
PF 12-FEB-2001; 2001US-0782378.
XX
PR 02-OCT-2000; 2000US-237747P.
XX
PA (UYNY) UNIV NEW YORK STATE RES FOUND.
XX
PI Hearing P, Bahou WF, Sandalon Z, Gnatenko DV;
XX
DR WPI; 2002-690619/74.
XX
PT Producing vector, by introducing vector having nucleotide sequence,
PT adenovirus inverted terminal repeats and packaging sequence, and
PT adeno-associated virus terminal repeat, into cell, and culturing cell
PT -
XX
PS Disclosure; Page 93-117; 191pp; English.
XX
CC The present invention relates to a new method of producing a vector. The
CC method involves introducing recombinant vector having nucleotide sequence
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'
CC end of NS, into cell expressing adenovirus early gene lacking from
CC vector; and culturing cell to produce another vector. The method is
CC useful for generating vectors, especially mAd vectors. The method is
CC useful in transferring nucleotide sequences of interest into a cell, for
CC gene transfer applications (e.g. gene therapy) in vitro, ex vivo and in
CC vivo. The nucleotide sequences are useful for treating diseases
CC associated with it, i.e. adenosine deaminase gene associated with
CC adenosine deaminase deficiency with severe combined immune deficiency,
CC beta-chain of haemoglobin gene associated with beta-thalassaemia and
CC sickle cell disease, receptor for low density lipoprotein gene associated

```
CC   with familial hypercholesterolaemia, hypoxanthine-guanine
CC   phosphoribosyltransferase associated with Lesch-Nyhan syndrome,
CC   phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,
CC   dystrophin gene associated with muscular dystrophy, and human cystic
CC   fibrosis transmembrane conductance regulator gene associated with cystic
CC   fibrosis. The present nucleic acid sequence represents a human disease
CC   gene sequence that was used in the methods of the invention.
XX
SQ   Sequence 56737 BP; 15689 A; 11281 C; 11599 G; 18168 T; 0 other;

  Query Match             3.9%;  Score 38.6;  DB 24;  Length 56737;
  Best Local Similarity   59.9%;  Pred. No. 41;
  Matches  100;  Conservative    0;  Mismatches   64;  Indels    3;  Gaps    2;

Qy        819 GGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCCATGTAGAATGAACCGTCCAGAGAG 878
              ||| |   | || ||||||||  ||||| |||||  |||||| ||||||    |    |
Db      22684 GGGGAGATGTAGCTCAGAGGATACAAAGTAGCAGATATGTAGGATGAACAAGTCTAGAAA 22625

Qy        879 CCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGCACTGTTCGTATTTGGAGTTCATGC 938
                  |     | | ||          | |        || | ||||||||  ||||  |
Db      22624 TATAATGTACAACATGAAGTATATAGGTAATAAAATTGTGCT-GTATTTGGGATTCACAC 22566

Qy        939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAA 985
              ||||||  |  ||| |||| |||||||||| |||| ||||||||||
Db      22565 TAAATGA--GATTTTAAGCTCCTCTTGCCACCAAAACAAAAAAAAAA 22521


RESULT 47
ABQ77405
ID   ABQ77405 standard; DNA; 185695 BP.
XX
AC   ABQ77405;
XX
DT   10-MAY-2003  (first entry)
XX
DE   Human THBS1 DNA.
XX
KW   Human; THBS1; vascular disease; cardiant; antiarteriosclerotic; stroke;
KW   cerebroprotective; gene therapy; coronary artery disease; ischaemia;
KW   myocardial infarction; peripheral vascular disease; pulmonary embolism;
KW   venous thromboembolism; forensic; paternity testing; GI12583762; gene;
KW   SNP; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace (53502,t)
FT                   /*tag=  a
FT                   /standard_name= "SNP"
FT                   /note= "Single nucleotide polymorphism (ID G334u3)
FT                          which does not change the THBS1 protein"
XX
PN   WO2003016494-A2.
XX
PD   27-FEB-2003.
XX
PF   16-AUG-2002; 2002WO-US26343.
XX
PR   16-AUG-2001; 2001US-313097P.
PR   05-OCT-2001; 2001US-327485P.
PR   14-DEC-2001; 2001US-0020141.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J,  Ableson A;
XX
DR   WPI; 2003-300617/29.
DR   P-PSDB; ABG74673.
XX
PT   Identifying a subject as a candidate for a particular course of therapy
PT   to treat a vascular disease or disorder, e.g. stroke, myocardial
PT   infarction or ischemia by determining the identity of the nucleotide
PT   present at specific positions  -
XX
PS   Claim 3; Fig 11; 568pp; English.
XX
CC   This invention describes a novel method for identifying a subject as a
CC   candidate for a particular course of therapy to treat a vascular disease
CC   or disorder. The method comprises determining the identity of the
CC   nucleotide present at specific positions, or their complements, and
CC   identifying the subject as a candidate for a particular clinical course
CC   of therapy based on the identity of the nucleotide present in that
CC   specific position. The method can be used for identifying a subject who
CC   is a candidate for further diagnostic evaluation of a vascular disease or
CC   disorder and selecting a clinical course of therapy. The products of the
CC   invention have cardiant, antiarteriosclerotic and cerebroprotective
CC   activity and can be used for gene therapy. The methods disclosed are
CC   useful for treating a vascular disease, e.g. atherosclerosis, coronary
CC   artery disease, myocardial infarction, ischaemia, stroke, peripheral
CC   vascular diseases, venous thromboembolism and pulmonary embolism. The DNA
CC   sequences are useful as fingerprint for detecting different individuals
CC   within the same species applicable in forensic studies and paternity
CC   testing. This sequence encodes the human THBS1 gene represented in
CC   GI12583762, used to illustrate the method of the invention.
XX
SQ   Sequence 185695 BP; 59388 A; 38636 C; 35664 G; 52007 T; 0 other;

  Query Match             3.9%;  Score 38.4;  DB 25;  Length 185695;
  Best Local Similarity   53.7%;  Pred. No. 68;
  Matches  101;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy        795 GAGGAGTTTGGGGGGTTCAGGATAGGGAATGGGGAGGTCAGAGGACGCAAAGCAGCAGCC 854
              || | | |   |  | |  | |  ||| |     |  | | ||||  |||||  |||| |
Db     112437 GATGTGGTGAGACTGGTAGGAAATGGGGAGATATACATAAAAGGATACAAAGTTGCAGAC 112496

Qy        855 ATGTAGAATGAACCGTCCAGAGAGCCAAGCACGGCAGAGGACTGCAGGCCATCAGCGTGC 914
              |||||  ||||||    | |   | | | ||   | | ||   |     |     | |
Db     112497 ATGTAAGATGAACAAGTCTGGAGGTCTAACATACAACATGAGGACTATAGTTAGTAATAC 112556

Qy        915 ACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAA 974
                ||  ||||||   ||| |||||||| ||||   |||||  || |||| ||||| | |
Db     112557 TGCCTT-TTATTTGAGATTCTTGCAAAAAGAGTAGATTTTACGTGTTCTTTCCACACATA 112615

Qy        975 AAAAAAAA 982
                |||  |
Db     112616 CGCAAAGA 112623


RESULT 48
ABV61090/c
ID   ABV61090 standard; cDNA; 294 BP.
XX
AC   ABV61090;
XX
DT   13-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 61081.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
```

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 11605-11606; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 294 BP; 104 A; 26 C; 69 G; 95 T; 0 other;

Query Match 3.8%; Score 38; DB 23; Length 294;
Best Local Similarity 81.5%; Pred. No. 9.9;
Matches 44; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 939 AAAATGAGTGTGTTTTAGCTGCTCTTGCCACAAAAAAAAAAAAAAAAAAAAAAAA 992
Db 93 AAAAGGTTTTTGTTTTACCCCCCCTTCCCCCAAAAAAAAAAAAAAAAAAAAAAAA 40

RESULT 49
ABQ81848
ID ABQ81848 standard; DNA; 349980 BP.
XX
AC ABQ81848;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1104.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS Bifidobacterium longum.
OS Synthetic.
XX
PN EP1227152-A1.
XX
PD 31-JUL-2002.
XX
PF 30-JAN-2001; 2001EP-0102050.
XX
PR 30-JAN-2001; 2001EP-0102050.
XX
PA (NEST) SOC PROD NESTLE SA.
XX
DR WPI; 2002-668397/72.
XX
PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample -
XX
PS Disclosure; SEQ ID 1104; 80pp; English.
XX
CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has antidiarrheic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification.
CC N.B. The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 69195 A; 106952 C; 106128 G; 67705 T; 0 other;

Query Match 3.8%; Score 38; DB 24; Length 349980;
Best Local Similarity 60.8%; Pred. No. 1e+02;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 255 ACCAAGAACATCAAGGTGGCCAAGAAGGTGGTGAAGACCCACGAGCCGGCCTCCTTCAAC 314
Db 222380 ACCAAGACCTTCAAGACCGACAAGAAGGTCCAGGCTTACTCCGAGGCCGGCTCCGTCATC 222439

Qy 315 CTCAACGTCACACTCAAGTCCAGTCCAGACCTGCTCACCTAC 356
Db 222440 CAGACCTTCACCATCCCGTCCTCCCTTGACCACTTCAAGAAC 222481

RESULT 50
ABV18588
ID ABV18588 standard; cDNA; 353 BP.
XX
AC ABV18588;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 18579.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 3060; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 353 BP; 129 A; 57 C; 66 G; 77 T; 24 other;

Query Match 3.8%; Score 37.8; DB 23; Length 353;
Best Local Similarity 63.5%; Pred. No. 12;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy        908 AGCGTGCACTGTTCGTATTTGGAGTTCATGCAAAATGAGTGTGTTTTAGCTGCTCTTGCC 967
              | |||  |||   ||||| | | | |  || | | | | | | | ||  |  ||  |    |
Db        268 ACCGTAAACTCACCGTATCTAGTGATTCTGTACATTAAATCTTTTGAACATGTTTNCCCN 327

Qy        968 ACAAAAAAAAAAAAAAAAAAAAAAA 992
              |||||||||||||||||||||||||
Db        328 NNAAAAAAAAAAAAAAAAAAAAAAA 352
```

Search completed: February 9, 2004, 13:11:14
Job time : 348 secs